

Universitetet i Oslo  
Institutt for geofag

# Flomrisikoanalyse for Hamar og Lillestrøm

Helge Bakkehøi

Candidatus  
Scientiarum

1. september 2003





---

## Abstract

This work focuses on the two towns most exposed for flooding in Norway: Lillestrøm which lies alongside Lake Øyeren, and Hamar which is situated near Norway's largest lake, Mjøsa. The main objective of this work is to produce damage curves for roads, railways and detached houses for these towns. However, in the Lillestrøm area, with fewer significant data, damage curves were produced **only** for roads and railways. A risk analyses was made by connecting a flood frequency analysis, geographical information and cost data from the big flood in 1995. Due to the lack of sufficient cost data, more focus was given to the uncertainties of the analyses rather than the actual results.

The results showed that the expected damage (risk) from flooding for all the detached houses in the Hamar region was 12 million NOK. The expected damage (risk) from flooding railways and roads are 37 375 NOK and 280 800 NOK, respectively. In Lillestrøm the expected damage (risk) from flooding railways and roads was equally to zero, due to the huge flood protection works in this area. The results from the uncertainty analyses showed that more work needs to be performed to better estimate the total risk.



# Innhold

<b><u>INNHold</u></b>	<b>4</b>
<b><u>1 INNLEDNING</u></b>	<b>6</b>
<u>1.1 BAKGRUNN</u>	6
<u>1.2 MÅL</u>	7
<u>1.3 OPPGAVENS UTFORMING</u>	7
<b><u>2 DATA</u></b>	<b>8</b>
<u>2.1 GEOGRAFISKE INFORMASJONSSYSTEMER (GIS)</u>	8
<u>2.1.1 Sosi-filer</u>	9
<u>2.1.2 GIS-verktøy</u>	10
<u>2.2 HYDROLOGISKE DATA</u>	10
<u>2.2.1 Mjøsa</u>	10
<u>2.2.2 Øyeren</u>	11
<u>2.3 SKADE</u>	12
<b><u>3 TEORI</u></b>	<b>14</b>
<u>3.1 LITT GENERELL STATISTIKK</u>	14
<u>3.1.1 Den uniforme fordelingen</u>	15
<u>3.1.2 Estimering av den kumulative fordelingsfunksjonen <math>F(x)</math></u>	15
<u>3.1.3 Kjernetetthetsestimering</u>	16
<u>3.2 FLOMFREKVENSPANALYSE</u>	16
<u>3.2.1 Bruk av L-momenter til Parameterestimering</u>	16
<u>3.2.2 GEV (Generalized Extreme Value)-fordelingen</u>	17
<u>3.2.3 Reduced Gumbel plott</u>	19
<u>3.2.4 Autokorrelasjonsplott</u>	20
<u>3.2.5 Gjentaksintervall</u>	21
<u>3.2.6 Trendtest</u>	22
<u>3.3 USIKKERHETSPANALYSE</u>	22
<u>3.3.1 Bootstrapping</u>	22
<u>3.3.2 Stokastisk simulering</u>	23
<u>3.3.3 Konfidensintervall</u>	24
<u>3.3.4 Konfidensbånd til den kumulative fordelingsfunksjonen</u>	25
<u>3.4 RISIKO</u>	26
<u>3.4.1 Skadefunksjonen</u>	26

<b><u>4</u></b>	<b><u>METODE</u></b>	<b>30</b>
4.1	<u>FLOMFREKVENSANALYSE</u>	30
4.2	<u>RISIKOBEREGNING</u>	31
4.3	<u>SKADEFUNKSJON FOR VEI/JERNBANE</u>	32
4.4	<u>SKADEFUNKSJONEN FOR BYGNINGER</u>	33
<b><u>5</u></b>	<b><u>RESULTATER OG DISKUSJON</u></b>	<b>34</b>
5.1	<u>INNLEDENDE FLOMFREKVENSANALYSE</u>	34
5.1.1	<u>Mjøsa</u>	35
5.1.2	<u>Øveren</u>	37
5.1.3	<u>Regional flomfrekvensanalyse</u>	39
5.2	<u>USIKKERHETSBEREGNINGER</u>	40
5.2.1	<u>Usikkerhet fra kostnadsberegninger</u>	40
5.2.2	<u>Usikkerheten fra flomfrekvensanalysen</u>	41
5.2.3	<u>Usikkerhet fra GIS-analysen</u>	44
5.2.4	<u>Generelt om usikkerhetsberegningene</u>	46
5.3	<u>RISIKOANALYSE</u>	46
5.3.1	<u>Lillestrøm</u>	46
5.3.2	<u>Hamar</u>	49
<b><u>6</u></b>	<b><u>KONKLUSJON</u></b>	<b>58</b>
	<b><u>TILLEGG A</u></b>	<b>60</b>
	<u>BYGNINGSFORDELING FOR HAMAR, KONSTRUERT PÅ BAKGRUNN AV SOSI-DATA</u>	60
	<b><u>TILLEGG B</u></b>	<b>64</b>
	<u>ARCINFO KOMMANDOER</u>	64
	<b><u>KILDELISTE</u></b>	<b>66</b>
	<b><u>STIKKORDREGISTER</u></b>	<b>68</b>

# 1 Innledning

## 1.1 Bakgrunn

Etter flommen på Østlandet i 1995 ble det satt i gang mye arbeid for å unngå at skadene blir like store ved framtidige flommer. Men det vil uansett kunne oppstå lignende situasjoner igjen, slik at beregninger av eventuelle skader fortsatt er aktuelt. Risikoanalyse er et relativt nytt felt innenfor hydrologien, men økende fordi det etter hvert har blitt tilgjengelig mye nyttig informasjon digitalt som kan behandles i geografiske informasjons systemer (GIS).

Det er ofte en del uklarheter i hva som ligger i begrepet risiko. I hydrologien opereres det med i hvert fall tre forskjellige definisjoner på risiko. I dette arbeidet vil risiko bli sett på som den forventede skaden, slik som det er definert i *bayesiansk statistikk* (Berger (1985)). Mer konkret for dette arbeidet vil risikoen bli sett på som gjennomsnittlig årlig skade angitt i norske kroner.

I dette arbeidet er det valgt å se på de to byene Lillestrøm og Hamar som grunnlag for risikoanalysen. Disse to byene ligger henholdsvis ved Øyeren og Mjøsa. Lillestrøm ble sterkt berørt av storflommene i 1966 og 67, etter dette ble det gjort mye arbeid med å unngå slike igjen, og skadene etter flommen i 1995 ble mye mindre enn en tilsvarende flom 30 år før. Likevel ga denne flommen skader for 100 millioner kroner. På bakgrunn av dette har Skedsmo kommune og NVE satt opp flomvoller og pumpestasjoner til 70 millioner kroner som beskytter hele byen mot i hvert fall en 300 års flom (106 m. o. h.).

Hamar har ikke blitt utsatt for flom på samme måte som Lillestrøm, men ble noe berørt under flommen i 1995 som var den sjette største registrerte flommen i Mjøsa siden målingene begynte på 1800-tallet. Fordi avrenningen fra Mjøsa er en av de to store tilførselene til Øyeren, kan en storflom i Øyeren bli unngått hvis man fyller opp Mjøsa. Det blir da et spørsmål hvilke av disse to byene som er viktigst, men det er et mye større potensielt skadeområde rundt Lillestrøm og sørover enn hva det er langs Mjøsa.

Oppgaven bygger på analyser gjort etter flommen i 1995. Det har ikke vært mulig å få tak i så mye data på skadetall, slik at det i dette arbeidet vil bli fokusert på skade på veier, jernbane og bolighus. Metodikken brukt i denne oppgaven bygger på *Gottschalk og Krasovskaia (1999)*. I tillegg er det innført noen usikkerhetsberegninger, samt skadefunksjon for veier og jernbane. Usikkerhetsberegningene er konsentrert rundt flomrisikoanalysen, da det er lite bakgrunnsdata for kostnader. All informasjon om kostnader er tatt fra *Wathne m. flere (1999)*. Det finnes mer informasjon om kostnader fra flommen i 1995 hos

forsikringsselskaper og takstmenn, men å få tak i denne informasjonen krever et omfattende arbeid.

## 1.2 Mål

Oppgavens mål er å undersøke skadene som kan oppstå under en eventuell flom ved to av de mest flomutsatte byene i Norge. Det vil bli gjort forsøk på å etablere skadekurver med Mjøsa og Øyeren som grunnlag. Det teoretiske vil få en sentral rolle i oppgaven, da det er begrenset med data tilgjengelig for å si noe om skade på annet enn veier, jernbaner og bolighus. Videre vil usikkerhetsanalysen bli viktig, denne vil bli utført på bakgrunn av bootstrapping. Det vil da være usikkerheten fra den innledende flomfrekvensanalysen som vil stå mest sentralt. Det vil bli utarbeidet skadekurver for de bygningstyper der kostnadsdata er lett tilgjengelig. Veier og jernbane vil også bli sett på, i tillegg utarbeides skadekurver for disse. Det vil bli implementert usikkerhet i alle skadekurvene.

Tilsvarende analyser gjort etter flommen i 1995 bygger på vannføringsdata, mens det i dette arbeidet blir brukt vannstandsdata fra Mjøsa og Øyeren.

## 1.3 Oppgavens utforming

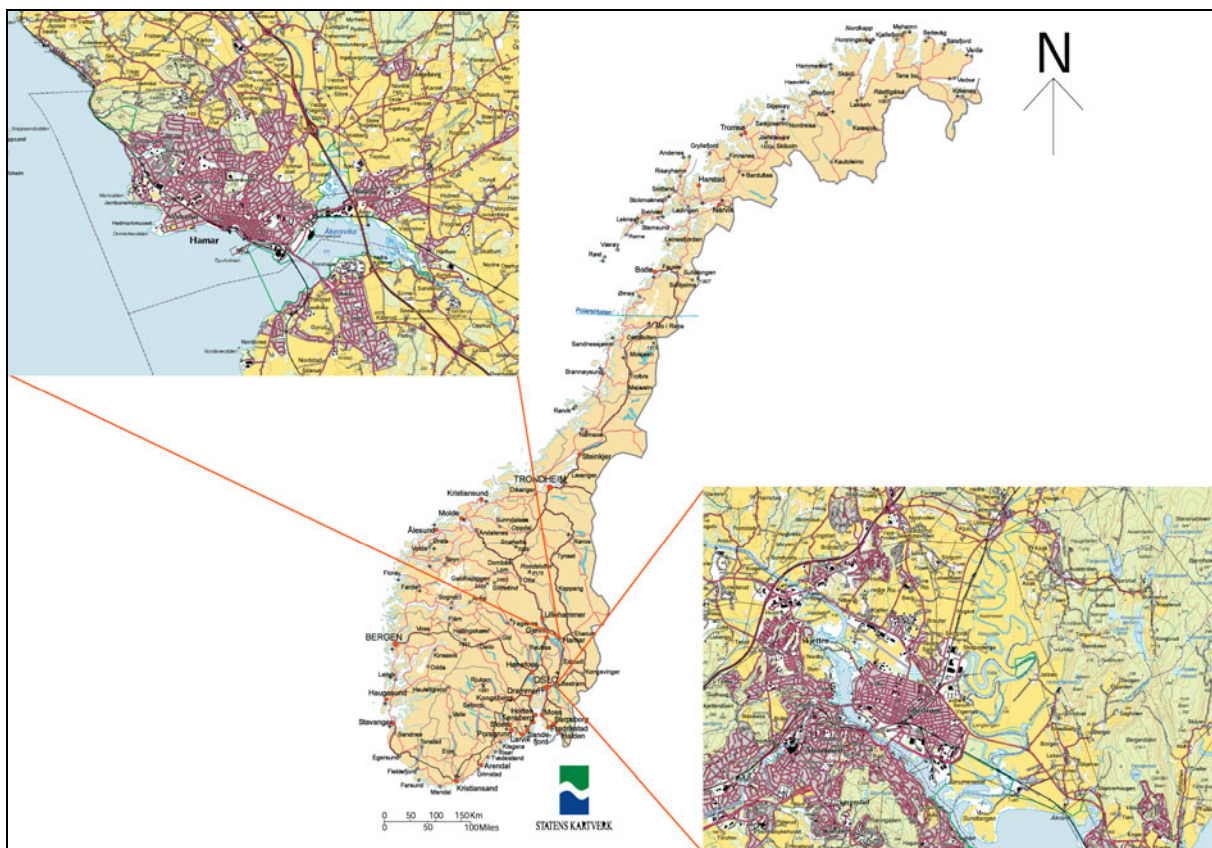
Oppgaven er delt inn i 6. kapitler. Kapittel 2 vil ta for seg de data som er brukt i oppgaven. Der vil de tre hovedkomponenten i oppgaven bli presentert, som er vannstandsdata, skadetall og GIS-data. Videre vil det i kapittel 3 bli presentert den statistiske bakgrunnsteorien til dette arbeidet, og hvordan dette blir brukt i oppgaven vil bli presentert under kapittel 4 som blir kalt metode. Resultatene blir først presentert i kapittel 5, som er delt inn i 3 underkapitler. Det er innledende analyse, usikkerhetsberegninger og til slutt risikoanalyse. En oppsummering kommer under konklusjonen i kapittel 6. Til slutt er det to vedleggs kapitler, hvor det første er en tabell over berørte bygninger under tidligere storflommer og det andre er en oversikt over kommandoer som er brukt i GIS-verktøyet Arcinfo.



## 2 Data

### 2.1 Geografiske informasjonssystemer (GIS)

For å kunne gjennomføre en flomrisikoanalyse kreves det en detaljert informasjon om beliggenheten til de eventuelt oversvømte objektene. Til dette er det brukt kartdata som enten er oppmålt av kommunene eller Statens kartverk, og deretter digitalisert. Disse dataene er samlet inn igjennom Skedsmo kommune og NVE. Dataene fra NVE er brukt i analysen for Hamar. Det var ikke mulig å få tak i nyere kartgrunnlag enn fra 1999. For Lillestrøm fantes det data, men dataene var lite differensierte, slik at det ikke var mulig å skille mellom forskjellige typer hus. Derfor vil det her kun bli utarbeidet en skadekurve for veier og jernbane. Disse dataene er tilveiebrakt fra Skedsmo kommune. Kartdataene for både Hamar og Lillestrøm har NGO akse III som koordinatsystem. Andre kart brukt i oppgaven kommer fra Statens kartverk.



*Figur 1: Oversiktskart over områdene sett på i dette arbeidet. Kartene er fra Statens kartverk, N50 serie.*

### 2.1.1 Sosi-filer

Sosi-standarden er en norsk standard for å presentere kartdata digitalt, og står for ”Samordnet opplegg for stedfestet informasjon”. Den ble første gang introdusert i 1987, og har senere blitt forandret for å tilpasse seg en mer internasjonal standard. Sosi-filer deles ofte opp i objekter som kommer i forskjellige filer. De objektene som er brukt i oppgaven er:

- TERR - Høydeinformasjon
- BYGG - Bygninger
- VANN - Innsjøer og vassdrag
- VSIT - Veisituasjon
- BANE - Jernbanedata

I en sosi-fil ligger det flere egenskaper til dataene. Alle typer karttegn har hver sin temakode, slik at det går greit å skille de forskjellige elementene fra hverandre. Hvis vi for eksempel tar for oss veier, så har veikant temakode 7002, mens veiens midtlinjen har 7003. Når det gjelder bygninger, så er ofte disse både presentert som flater og linjer. Flater til bygninger har temakode 5001.

Objekter har ofte egne unike elementer. Bygninger har et element som står meget sentralt i oppgaven, det er **BYGGTYP\_NBR**. Denne gir informasjon om hva slags type bygning vi har med å gjøre. De fleste digitale kart inneholder denne informasjonen, og det gjør arbeidet med en flomrisikoanalyse mye enklere. Hvis man da vet skadefunksjonen for de forskjellige boligtypene, så går det nesten automatisk å regne ut totalskaden.

Hvis denne informasjonen derimot ikke er implementert, må man synfare området og samle inn denne informasjonen på egenhånd, før man kan kjøre en flomrisikoanalyse. En alternativ løsning til denne meget tidkrevende prosessen er å anta at to bygninger med omtrent like stort areal påføres omtrent den samme skaden. Arealet er ofte implementert i sosi-dataene, men det er uansett greit å regne ut i et GIS-verktøy (for eksempel Arcinfo).

Type Bygg	BYGGTYP_NBR
Enebolig	111
Garasje, uthus anneks til bolig	181
Lagerhall	231
Verkstedbygning	212

*Tabell 2.1.1: Tabellen viser BYGGTYP\_NBR for noen av de mest flomutsatte bygningstypene i Hamar kommune, en total oversikt finnes i vedlegg A.*

### 2.1.2 GIS-verktøy

Sosi-filene har blitt konvertert, og videre bearbeidet i GIS-verktøyet *Arcinfo 8.1*. Programmet Arcview har blitt brukt til enkel visualisering. Dataene har videre blitt eksportert til statistikkpakken R, som er et kraftig, gratis statistikkprogram. R er en klon av det kommersielle programmet S-plus, men det går mye raskere å regne ut for-løkker her! S-plus har også blitt brukt til enkelte enkle beregninger.

## 2.2 Hydrologiske data

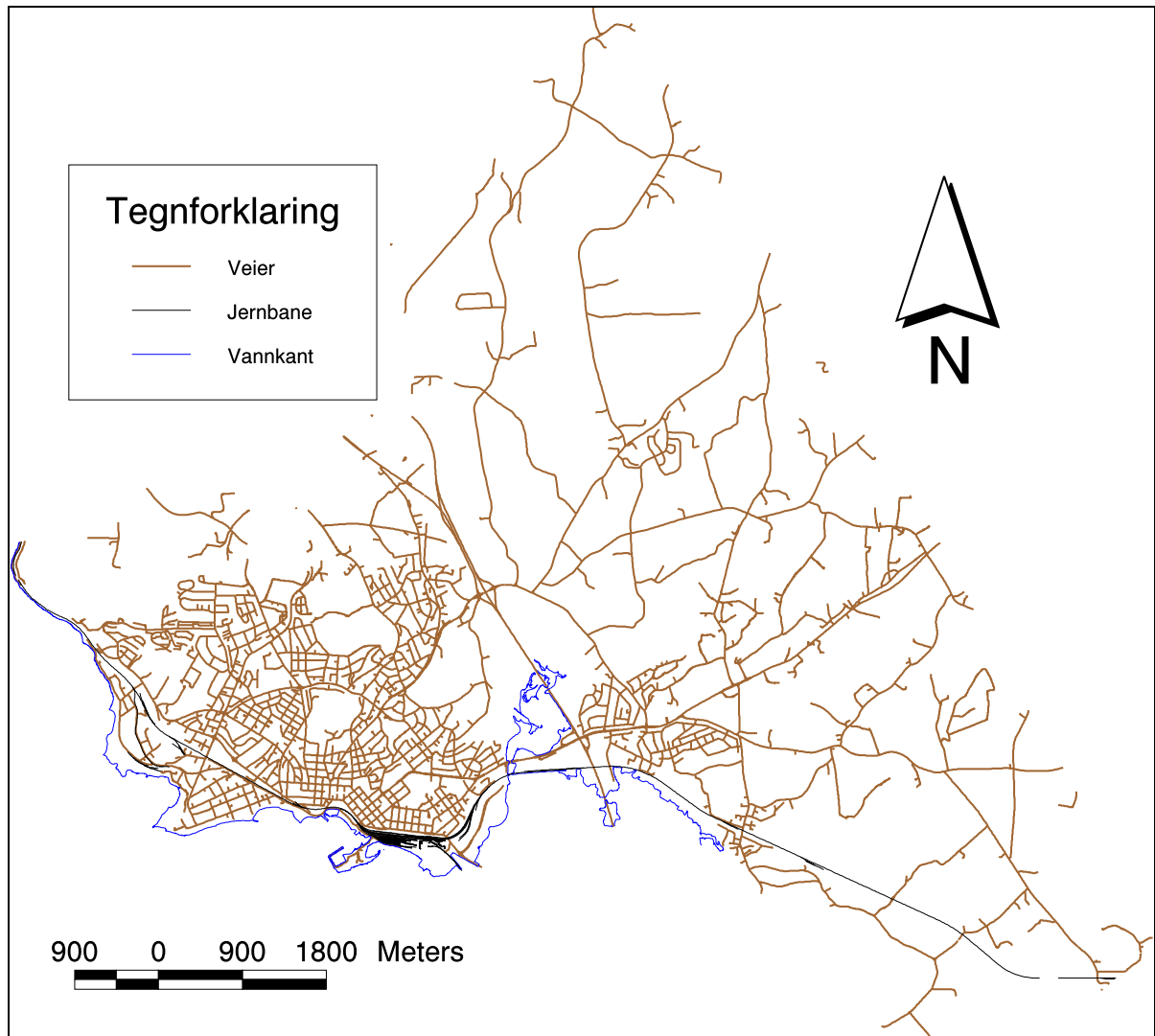
Vannstandsdata som er brukt i oppgaven er fra NVE, og går fra 1908-2000 for Mjøsa og 1881-1997 for Øyeren. Målestasjonene som er brukt er for Øyeren 2.125 *Mørkfoss* og for Mjøsa 2.10 *Hamar*. Vannstandsdataene er presentert som daglige maksimumsverdier.

Når vannstanden i Mjøsa og Øyeren er 0, ligger vannet henholdsvis 117,694 og 96,538 m. o. h. på bakgrunn av Statens kartverks høydesystem NN1954.

### 2.2.1 Mjøsa

Mjøsa er Norges største innsjø, og har Gudbrandsdalslågen som sin viktigste tilførselselv. I løpet av 1800 og 1900-tallet har Mjøsa blitt regulert 5 ganger. Det var i 1854, 1907, 1940, 1947 og 1961. Det ble i 1854 oppført en demning sør for Eidsvoll, denne ble revet i 1907. I 1909 ble det satt opp en ny demning litt lenger ned i Vormå, ved Svanfoss. Her er det gjort forandringer utover på 1900-tallet, de siste store reguleringene ble gjort i 1961. Reguleringene har ført til høyere vannstand på høst og vinter og lavere vannstand resten av året, mens flomvannstanden har blitt lavere. Oppstrøms er det gjort en god del reguleringer i det forrige århundre i forbindelse

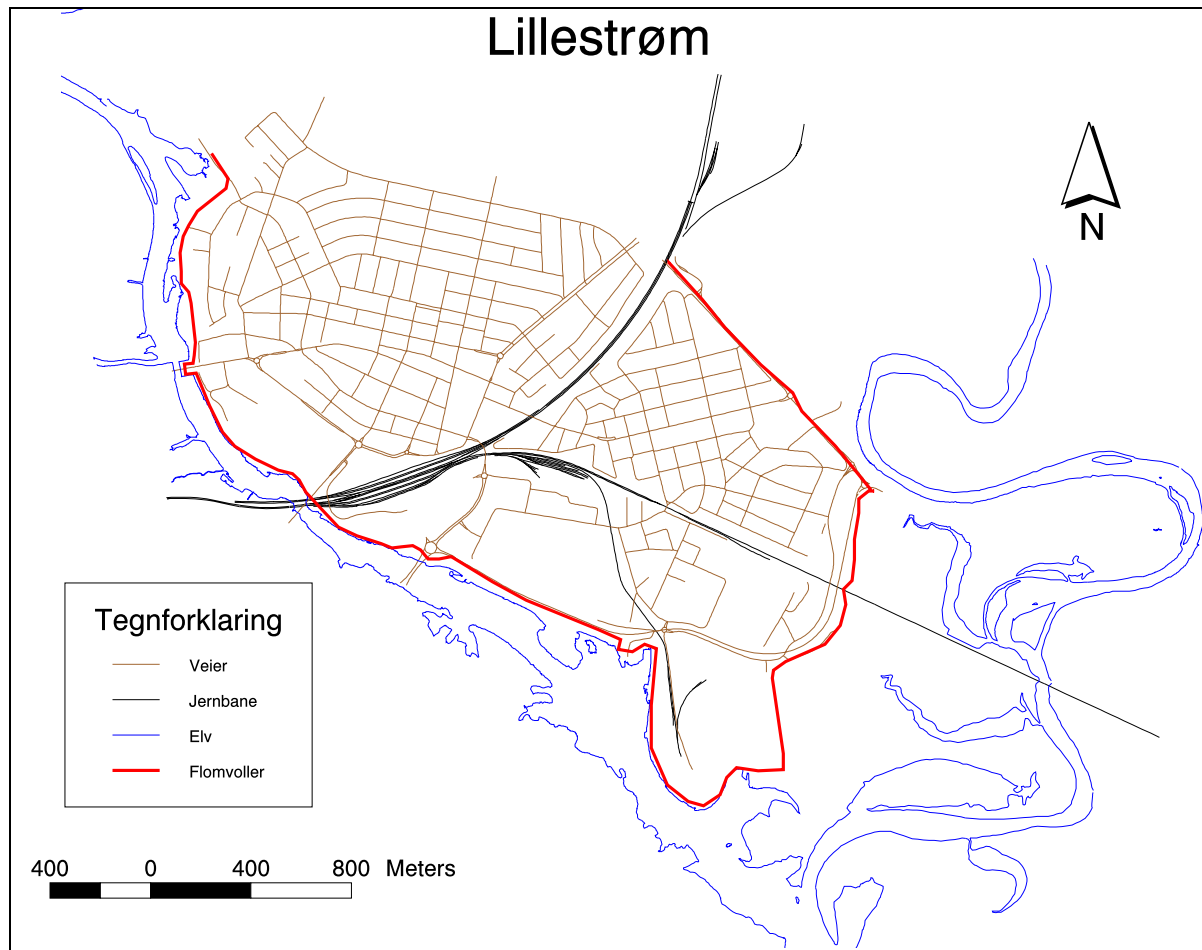
med kraftverksutbyggingen, men ingen større vassdrag er regulert etter Aursjø i 1962. Det vil si at det er gjort minimalt med regulering etter 1962, og det er naturlig at vassdraget i framtiden oppfører seg nokså likt som i perioden fra 1962 og fram til i dag.



*Figur 2: Kartet over viser veier og jernbane brukt i analysen for Hamar.*

### 2.2.2 Øyeren

Øyeren er Norges niende største innsjø, og får sin tilførsel først og fremst fra Glomma og Vorma. Lillestrøm ligger nord for tilløpet fra Glomma i en bukt kalt Svellet, her får Øyeren tilførsel fra to mindre elver, Leira og Nittelva. Leira og Nittelva drenerer store deler av Romerike, og Nittelva drenerer også betydelige områder av Nord- og Østmarka. Utløpet av Øyeren ligger ved Mørkfoss, hvor Glomma fortsetter ned mot Fredrikstad.



Figur 3: Kart over de veiene og jernbanene i Lillestrøm som er analysert.

Det har vært gjort noen forsøk på senkningsarbeid av Øyeren på midten av 1800-tallet, men det var først etter at Solbergfoss kraftverk ble bygget på 1920-tallet at det ble noen forbedringer. Flommen i 1966 og 1967 gjorde stor skade i Lillestrøm, og det ble satt i gang ytterligere flomreduserende tiltak, som ble ferdig i 1974. Etter flommen i 1995 er bare driftsikkerheten i kraftverket blitt utbedret. I dag skjer reguleringen av Øyeren ved Solbergfoss kraftverk, 5 km sør for Øyerens utløp.

Oppstrøms får Øyeren sin tilførsel både fra Lågen- og Glommavassdraget, men det er kun gjort små reguleringer her etter siste regulering av Øyeren. Det vil si at vannføringen har stabilisert seg etter 1973.

## 2.3 Skade

Som grunnlag for beregningen av kostnader er det brukt tall utarbeidet av *Wathne, m.flere(1999)*. Når det gjelder bygninger har det kun blitt gjort analyser for bolighus,

da dette er den eneste bygningstypen det finnes data lett tilgjengelig for. Det er brukt følgende regresjonslikning som standard for å beregne individuelle boligkostnader:

$$\text{''skade i 1000 kr''} = 1,2405 * \text{''vannhøyde fra grunnmur(cm)''}$$

(1)

Likning 1 er funnet ved å plote skaden på hvert hus etter 1995-flommen mot hvor høyt vannstanden kom på de forskjellige husene, og tilpasset dette med en lineær regresjonsanalyse.

Det er antatt at et bolighus blir påført skade fra og med når vannet er i kontakt med grunnmuren. 2,5 meter er satt som et standardmål på grunnmurens høyde.

For vei og jernbane er det brukt følgende tall:

- Vei: 360.000 kr/km
- Jernbane: 325.000 kr/km

Prisene over er tatt fra *Wathne, m. flere(1999)*.

Alle kostnadsberegninger som er brukt i denne analysen bygger på data som ble samlet inn etter 1995-flommen. Det er ikke tatt hensyn til eventuelle prisstigninger i beregningene. Likevel kan det være greit å være klar over at det har vært en generell prisstigning i løpet av denne perioden, som blir beregnet som konsumprisindeksen. Konsumprisindeksen lå i 1995 på 94,2 %, nå ligger den på ca. 111%. Det vil si at prisøkningen fra 1995 til i dag er på ca. 17,8 %.

## 3 Teori

### 3.1 Litt generell statistikk

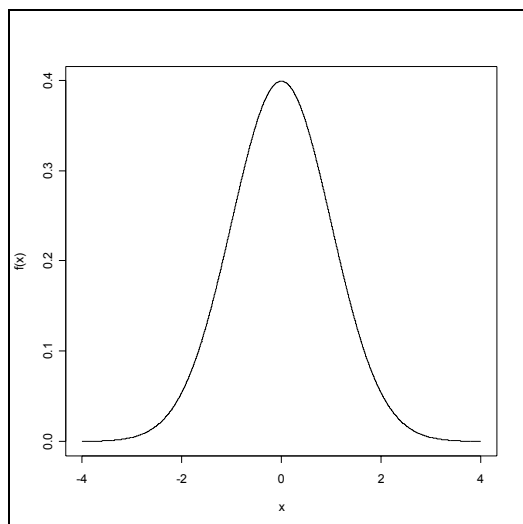
I statistikk opererer man med to forskjellig hovedmetoder å presentere data på:

1. *Sannsynlighetstettheten*, som sier noe om hvor de forskjellige dataen ligger i forhold til hverandre, på samme måte som et histogram.
2. *Kumulativ fordelingsfunksjon*, som sier noe om sannsynligheten for at data er lavere enn en x verdi.

#### 1. Sannsynlighetstetthet:

$$f(x) = \frac{dF(x)}{dx} \text{ der } \int_{-\infty}^{+\infty} f(x)dx = 1 \quad (2)$$

$$P(a < X < b) = \int_a^b f(x)dx \quad (3)$$

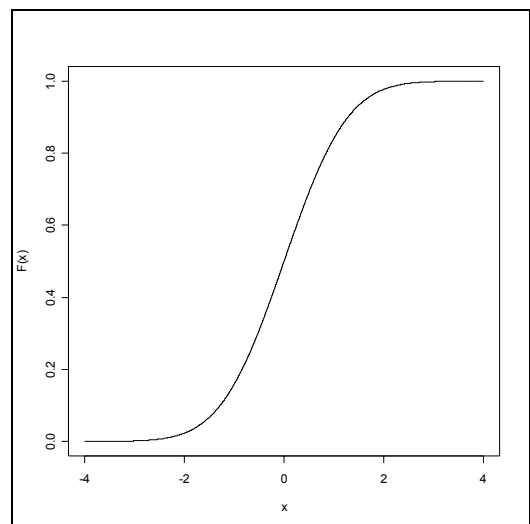


Figur 4: Standard normalfordeling presentert som en tetthetsfunksjon.

#### 2. Kumulativ fordelingsfunksjon:

$$F(x) = \int_{-\infty}^x f(t)dt \quad (4)$$

$$P(X < x) = F(x) \quad (5)$$



Figur 5: Standard normalfordeling presentert som en kumulativ fordelingsfunksjon.

$X$  er en tilfeldig variabel, og  $x$  er en gitt verdi. Det vil si at  $P(X < x)$  blir sannsynligheten for at en tilfeldig verdi i populasjonen er mindre enn verdien  $x$ .  $P(a < X < b)$  blir da sannsynligheten for at en tilfeldig variabel  $X$  ligger innenfor intervallet  $(a, b)$ .

Formlene beskrevet over er for teoretiske fordelinger, og krever først at man finner en tilpasning til de opprinnelige data. Det finnes også metoder for å finne fram til dette empirisk. Dette vil bli behandlet seinere. Noen vanlige fordelingsfunksjoner er normalfordelingen, lognormalfordelingen, og for ekstremverdier GEV-fordelingen. Metoder for å tilpasse slike fordelinger vil bli nærmere forklart i kapittel 3.2.1.

### 3.1.1 Den uniforme fordelingen

Den enkleste kontinuerlige fordelingsfunksjonen som er mulig er den uniforme fordelingen, ofte også kalt rektangelfordelingen:

$$f(x) = \frac{1}{b-a} \quad \text{når } a < x < b \quad (6)$$

$$E(x) = \frac{b+a}{2} \quad \text{Var}(X) = \frac{(b-a)^2}{12} \quad (7)$$

I en uniform fordeling vil  $f(x)$  være konstant for enhver  $x$ , altså i teorien skal alle verdier forekomme like mange ganger.

### 3.1.2 Estimering av den kumulative fordelingsfunksjonen $F(x)$ .

Et problem som ofte dukket opp i dette arbeidet, var at det var gitt en  $x$ -verdi, og så skal man finne  $F(x)$  fra den kumulative fordelingsfunksjonen. En grei måte å løse dette problemet på, er å finne antall verdier som er mindre enn  $x$ , for så å dele på antall verdier totalt, matematisk blir det slik:

$$F_X(x) = P_X(X < x) = \frac{1}{n} \sum_{i=1}^n I\{X_n < x\}, \quad (8)$$

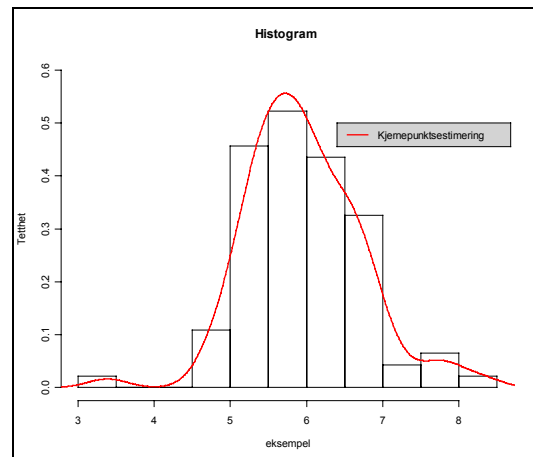
der  $I\{X < x\}$  er indikatorfunksjonen som er definert som  $I\{X_n < x\} = \begin{cases} 1 & \text{hvis } (X_n < x) \\ 0 & \text{hvis } (X_n > x) \end{cases}$



### 3.1.3 Kjernetetthetsestimering

For å sammenligne opprinnelige data med en fordelingsfunksjon, er det vanlig å bruke et histogram. Men dette blir ofte nokså unøyaktig, og det kan være greit å bruke en mer avansert metode.

Kjernetetthetsestimering er en metode for å estimere  $f$  basert på data  $x_1, \dots, x_n$  uif  $\sim f$ , og det er ganske tung teori som ligger bak som ikke tas med her. Nærmere info finnes i *Venables og Ripley(1999)*.



Figur 6: Kjernetetthetsestimering sammenlignet med histogram

## 3.2 Flomfrekvensanalyse

Flomfrekvensanalyse er mye brukt i hydrologien til å finne gjentakintervall, og vil i dette arbeidet ha en sentral posisjon som bakgrunn for risikoberegningene. I en flomfrekvensanalyse plukker man ut ekstremverdiene fra vannføringsseriene, og så tilpasses disse en fordeling ved hjelp av en esitmeringsmetode (beskrevet i kapittel 3.2.1). Det finnes to hovedmetoder å plukke ut ekstremverdiene på. Man kan enten plukke ut de høyeste årlige verdiene (AM), eller plukke ut verdier over en viss terskel (POT). Dette vil bli nærmere beskrevet i kapittel 4.1. To viktige antagelser i flomfrekvensanalysen er at data er uavhengig identiske fordelte variable (uif) og at de ikke forandres med tiden (stasjonære). I tillegg er det viktig at alle data kommer fra den samme fordelingen, altså at data er homogene. Et autokorrelasjonsplot kan brukes til å si noe om uavhengighet til dataene, mens en trendtest (lineær regresjon) kan brukes til å kontrollere for stasjonærhet.

Hvis det er flere vannstandsmålinger tilgjengelig i det samme område er det en fordel å normalisere data og kjøre en regional flomfrekvensanalyse. Da vil datagrunnlaget bli større og man vil oppnå en mer robust analyse. Dette krever homogenitet i data. Mer informasjon om dette kan finnes i *Gottschalk og Krasovskaia (2001)*.

### 3.2.1 Bruk av L-momenter til Parameterestimering

L-momenter er en metode som brukes til å estimere parametere i en teoretisk fordeling slik at den blir best mulig tilpasset til data. L-momenter ble første gang

introdusert av *Hosking (1990)*, og er definert som lineære kombinasjoner av data, ordnet i stigende rekkefølge. Disse har de senere år begynt å erstatte bruken av ordinære momenter. Noe av grunnen til dette er at ordinære momenter gir stor vekt på outliers (data som avviker fra resten), og gir store variasjoner av parameterne for små datasett. L-momenter fungerer bra til parameterestimering av fordelinger med lange haler, for eksempel ekstremverdifordelinger. L-momenter er utarbeidet av hydrologer, men er nå også på vei inn i den ordinære statistikken.

Ordinære momenter er definert slik:  $E(X^n)$ . Forventningen  $E(X)$  er første ordens moment. I tillegg er det noe som heter sentralmoment som er definert slik:  $E[(X - m)^n]$ . Andre ordens sentralmoment er definert som variansen, og tredje ordens sentralmoment er definert som skjevheten. Dette er verdier som både kan finnes teoretisk og empirisk, slik at det er mulig å estimere parameterne, ut i fra løsning av et likningssystem.

*Hosking m. flere* viste i 1985 at maximum likelihood estimering av små dataserier ga veldig ustabile svar, og anbefalt å bruke PWM (probability weighed moments) til estimeringen. L-momenter er en videreutvikling av PWM, som har større nøyaktighet og er enklere å bruke. L-momenter kan skrives som lineære kombinasjoner av PWM. For mer informasjon om maximum likelihood estimering se for eksempel *Larsen og Marx (1986)*

**L-moment er definert slik:**

$$\begin{aligned}\lambda_1 &= E(X_{1:1}) \\ \lambda_1 &= \frac{1}{2} E(X_{2:2} - X_{1:2}) \\ \lambda_3 &= \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \\ \lambda_4 &= \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})\end{aligned}\tag{9}$$

L-momentene sitt forholdstall, som er analogt til ordinære momenter er definert slik :

$$\tau_r = \lambda_r / \lambda_2, \quad r = 3, 4, \dots\tag{10}$$

$\tau_3$  er et mål på skjevhet og  $\tau_4$  er et mål på kurtosis.

### 3.2.2 GEV (Generalized Extreme Value)-fordelingen

GEV-fordelingen ble først introdusert av *Jenkinson (1955)*. Det er en 3-parameters generalisering av de tre ekstremverdifordelingene weibull ( $k > 0$ ), frechet ( $k < 0$ ) og gumbel ( $k \rightarrow 0$ ). GEV-fordelingen brukes i hydrologien til tilpasning av årlig maksimal vannstand. Den har også en tendens til å passe godt for andre typer

hydrologiske data, selv om dette ikke stemmer overens med teorien. For en dypere innføring i GEV-fordelingen henvises det til *Kotz og Nadarajah (2000)*.

Tetthetsfunksjonen til GEV-fordelingen er på denne formen:

$$f(x) = \frac{1}{a} \left[ 1 - k \left( \frac{x-u}{\alpha} \right) \right]^{1/k-1} e^{-\left[ 1 - k \left( \frac{x-u}{\alpha} \right) \right]^{1/k}} \quad (11)$$

der  $k$  er formparameter,  $u$  er lokasjonsparameter og  $\alpha$  er skalaparameter. Den kumulative funksjonen blir da slik:

$$F(x) = \exp \left\{ - \left[ 1 - \frac{k(x-u)}{\alpha} \right]^{1/k} \right\} \quad (12)$$

Det kan enkelte ganger også være interessant å finne den inverse til  $F$ ,  $F^{-1}(x)$ :

$$F^{-1}(x) = u + \frac{\alpha}{k} \left[ 1 - (-\ln(F))^k \right] \quad (13)$$

$F^{-1}(x)$  er ofte nyttig til simulering og for å kunne operere med kontinuerlige verdier.

Det har i denne oppgaven blitt brukt L-momenter som parameterestimeringsmetode, så de teoretiske L-momentene (definert i kapittel 3.2.1) til GEV-fordelingen er:

$$\lambda_1 = u + \frac{\alpha}{k} [1 - \Gamma(1+k)] \quad (14)$$

$$\lambda_2 = \frac{\alpha}{k} (1 - 2^{-k}) \Gamma(1+k) \quad (15)$$

$$\tau_3 = \frac{2(1 - 3^{-k})}{(1 - 2^{-k})} - 3 \quad (16)$$

Dette likningssystemet er det ikke mulig å løse, så vi må gjøre følgende tilnærming fra *Hosking m. flere (1985)*:

$$k \approx 7.8590c + 2.9554c^2, \quad (17)$$

$$c = \frac{2}{\tau_3 + 3} - \frac{\log 2}{\log 3} \quad (18)$$

De andre parameterne blir da som følger:

$$\alpha = \frac{\lambda_2 k}{(1 - 2^{-k}) \Gamma(1+k)}, \quad (19)$$

$$u = \lambda_1 - \frac{\alpha\{1 - \Gamma(1+k)\}}{k}, \quad (20)$$

der  $\Gamma$  er gammafunksjonen, som ligger innebygd i de fleste statistikkprogram.

### **Bakgrunn for GEV**

Bakgrunnen for at man som oftest bruker GEV-fordelingen når man tilpasser en fordelingsfunksjon til ekstremverdier, er analogt med det velkjente sentralgrenseteoremet, bare at man her bruker  $M_n = \text{Max}(X_1, \dots, X_n)$ , mens man i sentralgrenseteoremet bruker en sum  $S_n = (X_1 + \dots + X_n)$ .

**Teorem(Fisher og Tippett (1928)):**

Anta at  $(X_1, \dots, X_n)$  er uavhengige, identisk fordelte og tilfeldige variable med en fordelingsfunksjon  $F$ . Hvis det finnes en konstant  $a_n > 0$  og  $b_n \in \mathbb{R}$  slik at

$$\frac{M_n - b_n}{a_n} \rightarrow Y, \quad n \rightarrow \infty, \quad (21)$$

der  $Y$  er en tilfeldig variabel med fordelingsfunksjon  $G$ , så er  $G$  en av de tre ekstremverdifordelingene frechet, weibull eller gumbel.

### **3.2.3 Reduced Gumbel plott**

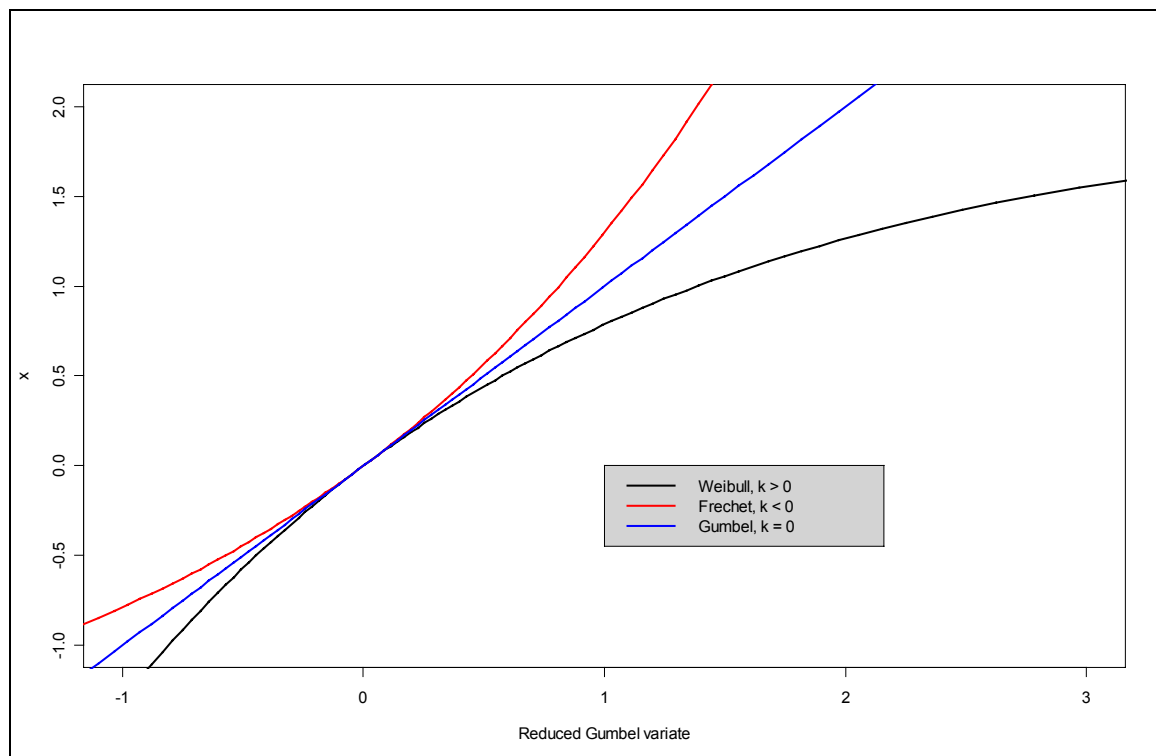
Dette er en måte å plote ekstremverdifordelinger på, introdusert av *Gumbel (1958)*. I et reduced gumbelplot, plotter man  $E[X] = -LN(-LN(F(x)))$  mot  $x$ . Blir kurven diagonal, har vi en gumbelfordeling.

For empirisk plotting av  $F(x)$  til data er det greit å bruke Gringortens-plotteformel:

$$F(x) = \frac{r - 0.44}{n + 0.12}, \quad \text{der } r = \{1, \dots, n\}, \text{ og } n \text{ er antall verdier.} \quad (22)$$

For å få mer forståelse av kurven er det greit å kombinere Gjentakintervallet med dette plottet. Sammenhengen mellom disse to blir fra *Gumbel (1958)* som følger:

$$-LN(-LN(F(x))) = \log(T) - \frac{1}{2T} \quad \text{der } T \text{ er gjentakintervallet.} \quad (23)$$



Figur 7: Figuren viser et reduced gumbel plot over GEV-fordelingen, med forskjellige verdier for formparameteren  $k$ .

Figur 7 viser at gumbelfordelingen i et slikt plott følger en rett linje, og weibullfordelingen konkav, mens frechetfordelingen er konveks. Av dette følger det da at weibull er oppad begrenset, mens frechet er nedad begrenset.

### 3.2.4 Autokorrelasjonsplott

Et autokorrelasjonsplott sier noe om avhengigheten mellom verdier over tid. Et slikt plott krever at det er konstant avstand mellom de målte data  $Y$ , for eksempel tidsserier. Hver posisjon i tidsserien blir identifisert med  $y_t$ , og  $y_{t+\tau}$  og er referert som et *lag* med lengde  $\tau$ .

Autokovariansen er definert slik:

$$\text{cov}_\tau = \frac{1}{n - \tau} \sum_{t=1+\tau}^n y_t y_{t-\tau} - \bar{Y} \bar{Y}_{t-\tau} \quad (24)$$

Autokorrelasjonen er normalisert autokovarians og blir da som følger:

$$r_{\tau} = \frac{\text{cov}_{\tau}}{\text{var } Y} = \frac{\left( \sum_{t=1+\tau}^n y_t y_{t-\tau} - (n-\tau) \bar{Y}_t \bar{Y}_{t-\tau} \right) / (n-\tau-1)}{\left( \sum_{t=1}^n (y_t - \bar{Y})^2 \right) / (n-1)} \quad (25)$$

En bredere innføring i dette temaet kan finnes i *Davis (2002)*.

### 3.2.5 Gjentaksintervall

Vi har den geometriske fordelingen som er definert slik:

$$p(y) = p(1-p)^y, y=1,2,\dots \quad (26)$$

$p$  er sannsynligheten for at en hendelse A skal inntreffe. Den tilfeldige variabelen  $Y$  er definert som antall ganger vi må gjenta forsøket før en hendelse A inntreffer. Forventningen til  $Y$  blir som følger:

$$E[Y] = \frac{1}{p} \quad (27)$$

Gjentaksintervallet kan nå finnes ved å innføre følgende:

- A: overstige vannstanden  $x$
- $F[X] = P(X < x)$ , som er fordelingsfunksjon for vannstanden  $X$ , som vil si sannsynlighet for at en tilfeldig vannstand  $X$  er mindre enn en gitt vannstand  $x$ .
- $p$  er sannsynligheten for at vannstanden  $x$  overstiges.
- $y$  er antall år det tar før  $x$  overstiges..

Videre har man følgende:

$$p = P(X > x) = 1 - P(X < x) = 1 - F[X] \quad (28)$$

Den geometriske fordelingen i dette tilfellet blir da:

$$p(y) = p(1-p)^y = (1-F(x))(1-(1-F(x)))^y, y=1,2,\dots \quad (29)$$

Det forventede antall år det tar før  $x$  er oversteget er da:

$$E[y] = \frac{1}{p} = \frac{1}{1-F(x)} = T(x) \quad (30)$$

Det er dette som i hydrologien brukes som definisjonen på gjentakintervallet, og  $F(x)$  finner man ut fra en flomfrekvensanalyse.

Med andre ord blir da gjentakintervallet gjennomsnittlig antall år det tar før en gitt flom blir oversteget. Hvis man har POT-data blir dette mer komplisert, og man må først konvertere data til AM. Mer info om dette finnes i *Gottschalk og Krasovskaia (2001)*

### 3.2.6 Trendtest

En trendtest er en test der det blir gjort en lineær regresjonsanalyse av data. En linje som følger denne likningen:  $y_i = a x_i + b$ , blir tilpasset data ved minste kvadraters metode. Hvis  $a$  er signifikant forskjellig fra 0, så kan man anta at det er trend i data, siden  $a$  er stigningstallet. Dette kan undersøkes ved og kjøre en såkalt t-test.

$$T = \frac{\bar{a}_n - 0}{s_n / \sqrt{n}}, \quad (31)$$

Da har vi at  $a$  er signifikant forskjellig fra null hvis T-verdien ligger utenfor følgende intervall:  $t_{(0.05)} < T < t_{(0.95)}$ . Der  $t_{( )}$  er  $(100 \cdot \cdot)$ -kvantilen til en standard  $t$ -fordeling.

## 3.3 Usikkerhetsanalyse

### 3.3.1 Bootstrapping

Bootstrapping ble introdusert i 1979 av Efron. Bootstrapping er en måte å finne usikkerheten til parameteren  $\theta$ , ut i fra simulering eller resampling. Metoden krever stor regnekraft, og har først blitt mulig å utnytte fullt ut i de senere åra.

Grunnideen bak bootstrapping er som følger: Vi kjenner ikke  $F$ , men la oss innføre et estimat på  $F$  kalt  $\hat{F}$ , og se på egenskapene til parameterne  $\hat{\theta}$  under  $\hat{F}$

Bakgrunnen for bruk av bootstrapping ligger i *store talls lov*:

Hvis  $X_1, X_2, X_3, \dots, X_n$  trekkes  $uif \sim G$  så gjelder følgende

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E[X] \quad \text{når } n \rightarrow \infty \quad (32)$$

Forventningsskjevheten, eller Bias som det ofte kalles, er definert slik:

$$\hat{Bias} = E^{\hat{F}}[\hat{\theta}] - \theta(\hat{F}) \quad \text{der} \quad E^{\hat{F}}[\hat{\theta}] = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i^* \quad (33)$$

Bias er altså ”gjennomsnittet av bootstrappsamplene” – ” $\theta$  estimatet”

Det er to mulige framgangsmåter i bootstrapp-verdenen, det er parametrisk og ikke-parametrisk bootstrapping. I parametrisk bootstrapping utfører man en simulering av fordelingsfunksjonen, mens i ikke-parametrisk bootstrapping resampler man data, med tilbakelegging.

Når vi har en korrekt parametrisk modell, kan det vises at ikke-parametrisk bootstrapping har dobbelt så stor varians som parametrisk bootstrapping. Til å kontrollere om en modell er korrekt, kan vi bruke en såkalt goodness-of-fit test, eller sammenligne kjernetetthetsestimatene til parameterne.

Simulering har utviklet seg til å bli et stort område innenfor statistikk, og det har blitt utarbeidet metoder for å simulere fra de fleste kjente fordelingsfunksjonene, også GEV-fordelingen, en mer detaljert beskrivelse finnes i kapittel 3.3.2.

Framgangsmåten for bootstrapping for GEV-fordelingen blir da som følger:

1. Simuler 10 000 datasampler fra GEV-fordelingen med inverteringsmetoden.
2. Finn parameterne ved hjelp av L-momenter til alle de 10 000 datasamplene.
3. Nå kan de 10 000 datasamplene brukes til beregning av f.eks standardavvik, bias og konfidensintervall til parameterne.

En annen metode som har blitt brukt til usikkerhetsestimering er jackknifing. Denne metoden er raskere, men ikke så god. Jackknifing brukes også til å lage konfidensintervall under  $BC_a$ -metoden. Mer informasjon om jackknifing finnes i Efron og Tibshirani(1998).

### 3.3.2 Stokastisk simulering

#### *Inverteringsmetoden:*

Inverteringsmetoden er en simuleringsmetode for inverterbare fordelingsfunksjoner. Det er en metode som er grei å implementere, og fungerer bra i de fleste situasjoner. De fleste funksjoner er inverterbare, men det er også mulig å tilnærme funksjonen slik at inverteringsmetoden fortsatt fungerer. GEV-fordelingen er en inverterbar funksjon.

Fremgangsmåten for å simulere fra GEV-fordelingen blir da som følger:

1. Generer data fra denne uniforme fordelingen,  $U \in [0,1]$



2. Sett så  $X = F^{-1}(U)$ , der  $F$  er den kumulative fordelingsfunksjonen for GEV.
3.  $X$ -ene vil da bli GEV-fordelte.

For mer informasjon om simuleringsmetoder se *Liestøl og Storvik (2001)*.

### *Resampling:*

Dette er en metode der man trekker tilfeldige data fra dataserien, og legger tallene tilbake etter hver gang. Metoden brukes mye innenfor bootstrappingsverdenen, og fungerer i de fleste situasjoner. Den brukes til ikke-parametrisk bootstrapping

### 3.3.3 Konfidensintervall

Når man skal finne konfidensintervallet til en parameter  $\theta$ , er man interessert i å finne A og B der  $P\{A \leq \theta \leq B\}$  er den ønskete dekningsgraden, for eksempel 95%.

Det finnes flere måter å finne konfidensintervaller fra bootstrapping på. Generelt har vi de klassiske metodene som normaltilnærmingen, og bruk av percentil-metoden. Disse metodene er ikke korrigert for skjevhet i fordelingen, og kan bli litt unøyaktige. En annen mer avansert metode er  $BC_a$ -metoden.  $BC_a$  metoden, ofte kalt ABC-metoden (Acceleration & Bias Corrected), er en andre ordens korrekt måte å finne konfidensintervallet på. Den er generelt sett på som en av de mest sikre metodene, og er definert slik:

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})}\right), \quad (34)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}\right), \quad (35)$$

her er  $\Phi(\cdot)$  standard kumulativ normalfordelingen med forventning 0 og standardavvik 1.  $z^{(\alpha)}$  er  $(100 * \alpha)$ -kvantilen til standard normalfordelingen.

Parameterne  $\hat{z}_0$  og  $\hat{a}$  er ikke helt greie å finne, men de kan finnes på blant annet følgende måte:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\text{antall}\{\hat{\theta} * (b) < \hat{\theta}\}}{B}\right), \text{ der } B \text{ er antall bootstrappsampler.} \quad (36)$$

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^3}{6 \left\{ \sum_{i=1}^n (\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)})^2 \right\}^{3/2}}, \quad \text{der } \hat{\theta}_{(i)} \text{ er jackknifesampler.} \quad (37)$$

$$\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \left( \frac{\hat{\theta}_{(i)}}{n} \right) \quad (38)$$

Konfidensintervallet blir da som følger:

$$KI = [\theta^{*(\alpha_1)}, \theta^{*(\alpha_2)}] \quad (39)$$

$\theta^{*(\cdot)}$  er  $(100*\cdot)$ -kvantilen til de simulerte parameterverdiene,  $\theta^*$

Mer informasjon om denne metoden finnes i *Efron og Tibshirani (1998)*.

### 3.3.4 Konfidensbånd til den kumulative fordelingsfunksjonen

Det finnes flere mulige fremgangsmåter for å finne et konfidensbånd til den kumulative fordelingsfunksjon.

De viktigste metodene er delt inn i globale og punktvis konfidensbånd.

- Punktvis bånd blir laget ved å plukke ut en god del  $x$ -verdier og finne konfidensintervallet til  $F(x)$ , for hver av disse  $x$ -verdiene. Denne trenger simulering.
- Globale bånd er samtidige bånd som er konstruert med hele fordelingen som grunnlag for alle verdier. Denne metoden kan klare seg uten simuleringer.

En tradisjonell fremgangsmåte for å utarbeide konfidensbånd er basert på Kolmogorov-Smirnov test. Den er global og ikke-parametrisk. Denne metoden har en tendens til å gi for stort konfidensintervall i halen, og båndet vil få en konstant bredde når  $F(x) \rightarrow 1$ . Dette passer dårlig i en flomfrekvensanalyse, der vi ofte er interessert i de høye  $F(x)$  verdiene. Selv om denne metoden er forholdsvis enkel å implementere, så er det bedre å bruke et punktvis konfidensbånd i denne oppgaven.

Andre forbedrede globale metoder hvor man for eksempel bootstrapper finnes, men de fleste bygger på maximum likelihood estimeringsmetode, se *Jeng og Meeker(2000)*

Et punktvis konfidensbånd er lettest å lage, og det konstrueres ved hjelp av bootstrapping. Det er viktig at dette konstrueres ut i fra  $F^{-1}(x)$ , da  $F(x)$  kan gi

undefinerte verdier. Man kjører 1000 simuleringer, og plukker ut øvre og nedre kvantil.

Ulempene med et punktvist konfidensbånd kontra et globalt, er at de statistiske egenskapene til fordelingen ikke blir bevart i hvert punkt. Et globalt konfidensbånd tar hensyn til dette.

Når man har konstruert et konfidensbånd til den kumulative fordelingsfunksjonen, er det rimelig greit å overføre dette til en skadekurve

## 3.4 Risiko

Fra Bayesiansk statistikk (*Berger (1985)*) er risiko definert som forventningen til en skadefunksjon,  $l(d(x), \theta)$ . Det kan da skrives slik:

$$R(d, \theta) = E_X[l(d(X), \theta)] = \int l(d(x), \theta) f_x(x | \theta) dx. \quad (40)$$

Poenget er å velge en handling  $a=d(X)$  fra et område  $A$  inneholdt alle mulige handlinger, basert på observasjoner av en tilfeldig variabel  $X$ .

Sannsynlighetsfordelingen til  $X$  er basert på en parameter  $\theta$ , som ofte blir kalt naturtilstanden.

Hvis vi skriver om risikoen til å passe med notasjonene i dette arbeidet, ser det ut som følger :

$$R(z, d(z)) = E_X[l(z, d(x))] = \int l(z, d(x)) f_x(x | z) dx \quad (41)$$

### 3.4.1 Skadefunksjonen

Det å finne skadefunksjonen (totale kostnader) er ganske rett fram ved å bruke elementær matematikk og statistikk:

*Den totale kostnaden = antall objekter oversvømt \* skade pr. objekt*

mer matematisk blir dette:  $l(y, z) = (n(y | z))l'(y)$  (42)

$l'(y)$  er skade for hvert enkelt objekt, og  $n(y | z)$  er antall objekter oversvømt.

Vannnivået i en bygning kan defineres slik,  $y = z-h$ , der  $h$  er høyden over havet til hvert enkelt objekt og  $z$  er en gitt vannstand(m.o.h). Den maksimale vannnivået man

da kan få blir  $y_{\max} = z - h_0$ , der  $h_0$  er høyde over havet til det lavest beliggende objektet.

$n(y | z)$  er da alle objekter som ligger innenfor  $y$  og  $y_{\max}$ , det er da det samme som:

*antall objekter \* sannsynligheten for at et objekt ligger innenfor  $y$  og  $y_{\max}$*

som matematisk kan skrives slik:

$$n(y_c | z) = N * P(y_c < Y < y_{\max}) = N \int_{y_c}^{y_{\max}} f_Y(y | z) dy \quad (43)$$

der  $N$  er totalt antall objekter.

Det vil da si at man er interessert i å finne  $f_Y(y | z)$  som er ukjent.

Det finnes i hvert fall to mulige fremgangsmåter å finne  $f_Y(y | z)$  på, den ene er kun å estimere  $f_Y(y | z)$  ved hjelp av kjernetetthetsestimeringsmetoder. Dette er en enkel tilnærmingmetode, og det blir da kurant å implementere konfidensintervall. Siden man er interessert i å finne en statistisk fremstilling av skadefunksjonen, så er den andre fremgangsmåten å foretrekke, som er en analytisk metode konstruert ved hjelp av elementær matematikk og statistikk utarbeidet av *Gottshalk og Kraskovskaia (1999)*:

Fra elementær statistikk har vi konvolveringen mellom  $f_Z$  og  $f_H$  som er definert slik:

$$f_Y(y) = \int_{-\infty}^{\infty} f_Z(z) f_H(y - z) dz \quad \text{når } y = z + h \quad (44)$$

For  $y = z - h$  kan man skrive om konvolveringsformelen litt:

$$f_Y(y) = \int_{-\infty}^{\infty} f_Z(z) f_H(z - y) dz \quad (45)$$

det kan videre skrives at  $f_Z(z)$  har en uniform fordeling:

$$f_Z(z) = \frac{1}{z_a - z_b} \quad (46)$$

der  $z_a$  og  $z_b$  er henholdsvis øvre og nedre grense for konfidensintervallet til vannstanden( $z$ ) ut i fra usikkerheten i høydemodellen. Den vil fungere som en glattingsparameter, der stor avstand mellom  $z_a$  og  $z_b$  gir en glattere kurve.

$$f_Y(y) = \int_{-\infty}^{\infty} f_Z(z) f_H(z - y) dz = \int_{-\infty}^{\infty} \left( \frac{1}{z_a - z_b} \right) f_H(z - y) dz = \left( \frac{1}{z_a - z_b} \right) \int_{-\infty}^{\infty} f_H(z - y) dz \quad (47)$$

Videre er definisjonen av den kumulative fordelingsfunksjonen (skrevet om for å passe inn i uttrykket over):

$$F_H(z_a - y) = \int_{-\infty}^{\infty} f_H(z - y) dz \quad (48)$$

Dette integralet kan da settes inn i likning 43, og man får følgende:

$$f_Y(y) = \frac{1}{z_a - z_b} F_H(z_a - y), \quad z_a - h_0 \leq y < z_b - h_0 \quad (49)$$

Dette kan igjen skrives slik:

$$f_Y(y) = \frac{1}{z_a - z_b} [F_H(z_a - y) - F_H(z_b - y)], \quad y < z_b - h_0 \quad (50)$$

Dermed er  $f_Y(y)$  beskrevet med et uttrykk man kan finne, og man kan da plotte  $f_Y(y)$  ved hjelp av likning 50. Plotting av en slik likning kan virke litt vrient, men det gjøres på følgende måte med å skrive om uttrykket for  $f_Y(y)$  litt:

$$f_Y(y) = \frac{1}{z_a - z_b} [F_H(z_a - y) - F_H(z_b - y)] = \frac{1}{z_a - z_b} \left[ \frac{1}{n} \sum_{i=1}^n I\{H < (z_a - y)\} - \frac{1}{n} \sum_{i=1}^n I\{H < (z_b - y)\} \right] \quad (51)$$

$$F_H(z_a - y) = P_H(H < (z_a - y)) = \frac{1}{n} \sum_{i=1}^n I\{H < (z_a - y)\} \quad (52)$$

$$F_H(z_b - y) = P_H(H < (z_b - y)) = \frac{1}{n} \sum_{i=1}^n I\{H < (z_b - y)\} \quad (53)$$

der  $I$  er en indikatorfunksjon som er 1 når innholdet er sant, og 0 når det er usant.

Når man nå har funnet  $f_Y(y)$  er man videre interessert i hvor mange objekter som blir oversvømt, det er gitt ved ligning 43.

Siden man ikke opererer med kontinuerlige data må vi skrive den slik:

$$\Delta n(y_c | z) = N \left[ f_Y(y_c | z) \left( \frac{y_{\max} - y_c}{m} \right) \right] \quad (54)$$

der  $m$  er antall  $y$ -verdier man har.

Nå har man alt man trenger for å beregne skadefunksjonen, som er slik:

$$l(y, z) = \Delta n(y | z) l'(y) \quad (55)$$

Vi er nå videre interessert i å finne  $f_L(l | y, z)$ , og et mye brukt teorem i statistikk sier at fra likning 55 så kan vi skrive  $f_L(l | y, z)$  slik:

$$f_L(l | y, z) = \frac{f_{L'}(l')}{\Delta n(y, z)} = \frac{f_{L'}(l(y, z)/\Delta n(y, z))}{\Delta n(y, z)} \quad (56)$$

Videre følger da forventningen til denne funksjonen:

$$E[L | y] = \int l \frac{f_{L'}(l(y, z)/\Delta n(y, z))}{\Delta n(y, z)} dy \quad (57)$$

Empirisk kan forventningen skrives slik:

$$E[L | y] = \bar{l} = \frac{1}{n} \sum_{i=1}^n [l | (y, z_i)] \quad (58)$$

Det er viktig når man skal plote disse data at man hele tiden passer på at man bruker de samme vannstandsdataene for skadefunksjonen og frekvensfordelingen.

## 4 Metode

### 4.1 Flomfrekvensanalyse

Flomfrekvensanalysen er en viktig del av risikoanalysen, og årsaken til dette vil bli beskrevet nærmere i kapittel 4.2. Det er alltid greiest å begynne med denne analysen, mye fordi dataene som ligger til grunn er forholdsvis lette å få tak i, og det er viktig å ta seg god tid med å studere vannstandsdataene før man setter i gang.

Det er viktig å anta at dataene er uavhengig før man går i gang med en flomfrekvensanalyse, da all bakgrunnsteori krever at data er uavhengige identiske fordelte variable.

I en flomfrekvensanalyse er man ikke interessert i alle de opprinnelige vannstandsdataene, men man plukker ut kun de ekstreme hendelsene og analyserer disse.

Dette kan gjøres på to forskjellige måter:

1. Plukke ut den høyeste verdien for hvert år, den såkalt AM-metoden.
2. Plukke ut alle verdier over en viss terskel, POT-metoden.

POT er en mer kompleks metode enn AM, og krever mer arbeid og analyse fra brukeren. Det kreves også i en POT-analyse at dataene er uavhengige av hverandre. Det kreves for så vidt også for AM-metoden, men siden man bare plukker ut en verdi per år så kan vi anta uavhengighet.

POT er sett på som en mer robust metode enn AM fordi den tillater mer data, så hvis det er mulig å bruke denne bør den brukes. AM kan i teorien gå glipp av den nest høyest verdien. I følge *Rao og Hamed (2000)* er AM mer statistisk effektiv enn POT når  $\lambda$  er liten ( $\lambda < 1.65$ ).  $\lambda$  er årlig gjennomsnittlig antall verdier som er over terskelen. Et annet problem med POT er at en verdi over terskelen ofte kommer dagen etter en annen som også er over terskelen. Det betyr at disse dataene er avhengige av hverandre, og den laveste må fjernes manuelt. Dette kan løses med å bruke en form for clusteranalyse hvor man kun tar vare på den største verdien innenfor et gitt intervall.

POT-data pleier å følge en ”generalized parreto” fordeling i motsetning til AM, som følger en GEV-fordeling. Valg av begge disse fordelingene samsvarer meget godt overens med teorien, slik at det ikke er noen grunn til å prøve å tilpasse data til andre fordelinger.

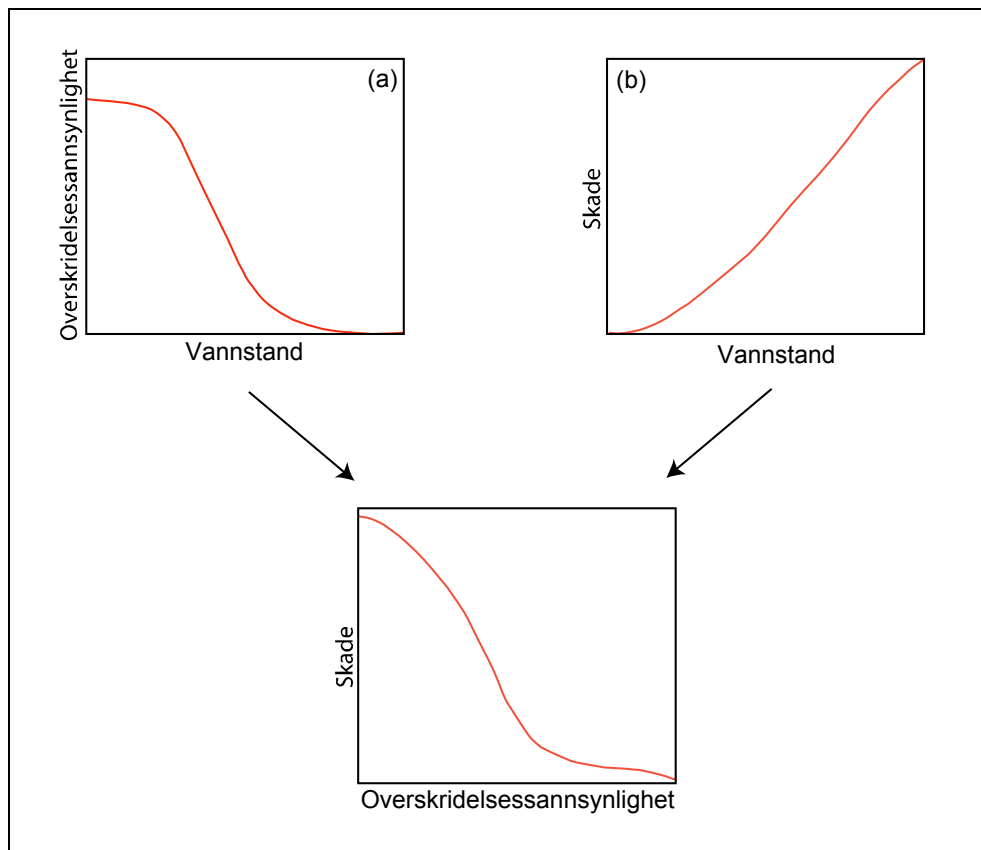
## 4.2 Risikoberegning

Flomrisikoanalysen tar utgangspunkt i to forskjellige analyser, en vanlig flomfrekvensanalyse og en skadefunksjonsanalyse. Disse to analysene kan til slutt kombineres og man får et plott som i Figur 25. En generell fremgangsmåte blir da å alltid ta utgangspunktet i et overskridelsessannsynlighetsplott og et vannstand-skadeplott som vist i Figur 8. Siden vannstanden er representert i begge de to plottene (a) og (b), kan da en skadekurve enkelt konstrueres ut i fra dette. Fra dette følger det at man kan skrive følgende:

$$F_z(z) = F_L(l(z)) = F_L(l) \quad (59)$$

$$r = \int_0^{\infty} l f_L(l) dl = \int_0^{\infty} l(z) f_z(z) dz \quad (60)$$

der  $l(z)$  er skadefunksjonen og  $f_z(z)$  er frekvensfordelingen til flommene. Risikoen  $r$  er forventningen til skadefunksjonen  $l$ , eller gjennomsnittlig årlig skade.



Figur 8: Figuren viser fremgangsmåten for å konstruere en skadekurve ved å kombinere et overskridelsessannsynlighetsplott (a) fra flomfrekvensanalysen og et vannstand-skadeplott (b) fra risikoanalysen.



Denne fremgangsmåten gjelder for alle skadeberegninger i dette arbeidet. Det er kun skadefunksjonen som ligger til grunn for plott (b) som varierer om det er vei, jernbaner eller bolighus man studerer.

### 4.3 Skadefunksjon for vei/jernbane.

For veier og jernbane er det blitt brukt en skadefunksjon som ikke bygger på statistiske beregninger. Metodene brukt for disse analysene er enkle og rett fram. Den er helt forskjellig fra skadefunksjonen presentert i 4.2.

#### **Framgangsmåten for å finne den total kostnaden blir da:**

1. Plukke ut midtlinjen til veiene fra en sosi-fil. Dette er kanskje ikke implementert i den nyeste sosi-standard.
2. Konverter disse til gridceller på for eksempel 5x5 meter, eller helst så lite som mulig
3. Tilordne høydeverdier til gridet ut i fra en høydemodell
4. Plukke ut en god del aktuelle høydeverdier, og tell opp antall gridceller som ligger lavere enn disse høydeverdiene.
5. Nå kan disse høydeverdiene multipliseres med gridstørrelsen, og vi får antall kilometer oversvømt for gitte vannstander.
6. Siden vi vet kostnaden pr. kilometer har vi nok til å etablere en vannstand-skadekurve for vei eller jernbane.
7. I tillegg trenger man å konstruerer en overskridelsessannsynlighetskurve fra flomfrekvensanalysen.
8. Skadekurven kan nå konstrueres ut i fra overskridelsessannsynlighetsplottet og vannstand-skadeplottet slik som Figur 8 viser.

Dette er en veldig enkel fremgangsmåte, og det er her blitt brukt en god del tilnærminger. Det er ikke tatt hensyn til at skaden blir større ved en langvarig flom, og det er kun brukt et kostnadskille: oversvømt eller ikke-oversvømt.

Det kan også oppstå skade, selv om flommen ikke når helt opp til objektet. For jernbane er det foreslått av *Wathne m. flere(1999)* at man skal dele opp i 3 lag: Formasjonsplanet, overbygning og skinner. Skadene vil da begynne cirka en meter under bakken, og øke med høyden.

En flom som er så stor at den gir store flomskader i Lillestrøm og Hamar, er ofte forenlig med at små sideelver også går over sine bredder og forsterker flommen. Det er ofte slike sideelver som gjør de største skadene på veier og jernbaner.

I kostnadsdataene presentert i kapittel 2.3 er den totale skaden tatt med , inklusiv skaden fra sideelver. Disse kostnadene vil da trolig bli litt for høye i forhold til det reelle tallet. Det er jo veier/jernbane oversvømt av innsjøen som er interessant i oppgaven.

## 4.4 Skadefunksjonen for bygninger

For å finne skadefunksjonen til bygninger, er det brukt en mer detaljert beregning enn den for vei/jernbane. Her er det fulgt teorien fra kapittel 3.4.1.

Begrensede mengder data medførte at bolighus var den eneste bygningsformen som ble sett på i oppgaven. Men det er viktig å merke seg at akkurat den samme metoden kan brukes til hvilke som helst annen bygningstype.

### **En enkel beskrivelse på hvordan dette kan gjøres i praksis:**

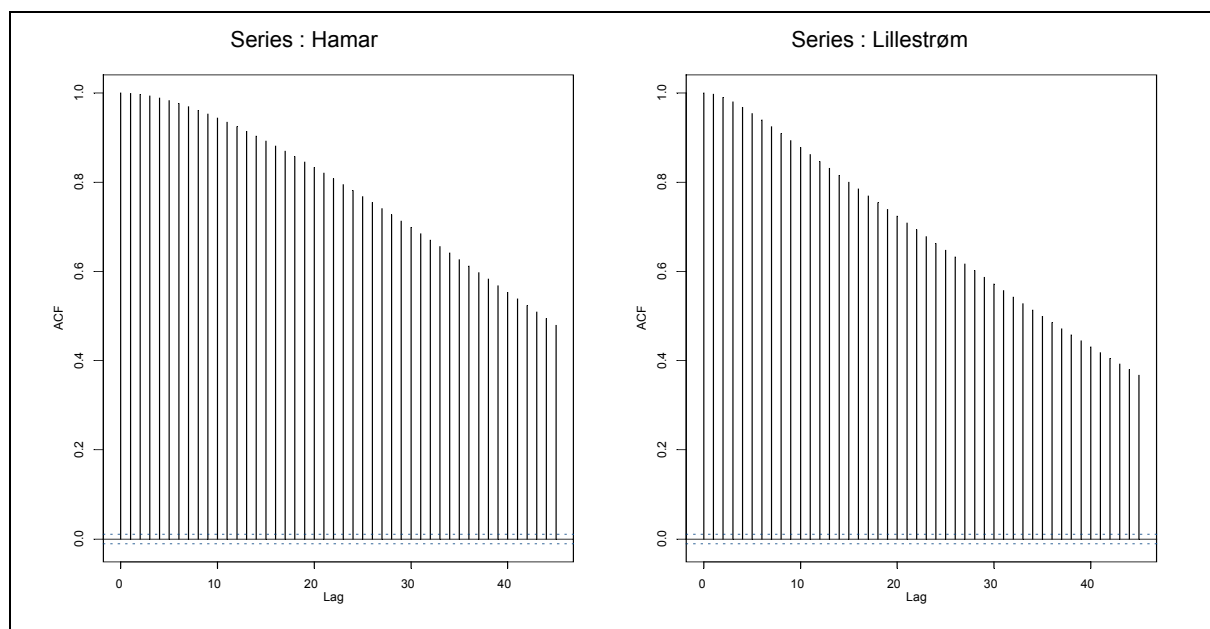
1. Gjør en flomfrekvensanalyse for det aktuelle området. En regional flomfrekvensanalyse er å foretrekke hvis det er oppnåelig.
2. Importer Sosi-filen i Arcinfo
3. Plukk ut alle flater med temakode 5001 (bygninger) og kun de typer bygninger hvor skadefunksjonen er kjent, fra en sosi-fil.
4. Tilordne disse flatene høydeverdier ved hjelp av en terrengmodell
5. Eksporter dette inn i et statistikkprogram, og gjør resten av analysen der.
6. For så videre å kunne lage en vannstand-skadekurve kan man følge metoden for skadefunksjonen beskrevet i kapittel 3.4.1.
7. I tillegg trenger man å konstruerer en overskridelsessannsynlighetskurve fra flomfrekvensanalysen.
8. Skadekurven kan nå konstrueres ut i fra overskridelsessannsynlighetsplottet og vannstand-skadeplottet slik som Figur 8 viser.

Terrengmodellen er i oppgaven konstruert ut i fra høydekurver med 1 og 5 meters ekvidistanse. Interpolasjonsmetoden som er brukt til dette er topogrid. Nærmere beskrivelse av denne metoden kan finnes i *Hutchinson(1989)*.

## 5 Resultater og diskusjon

### 5.1 Innledende flomfrekvensanalyse

Som bakgrunn for risikoanalysen kreves det at det gjøres en flomfrekvensanalyse. En viktig forutsetning for å kunne utføre en slik er at data er uavhengige identiske fordelte variable. Det er ofte vanskelig å få data som tilfredsstiller dette kravet, så det kreves en tilnærming.

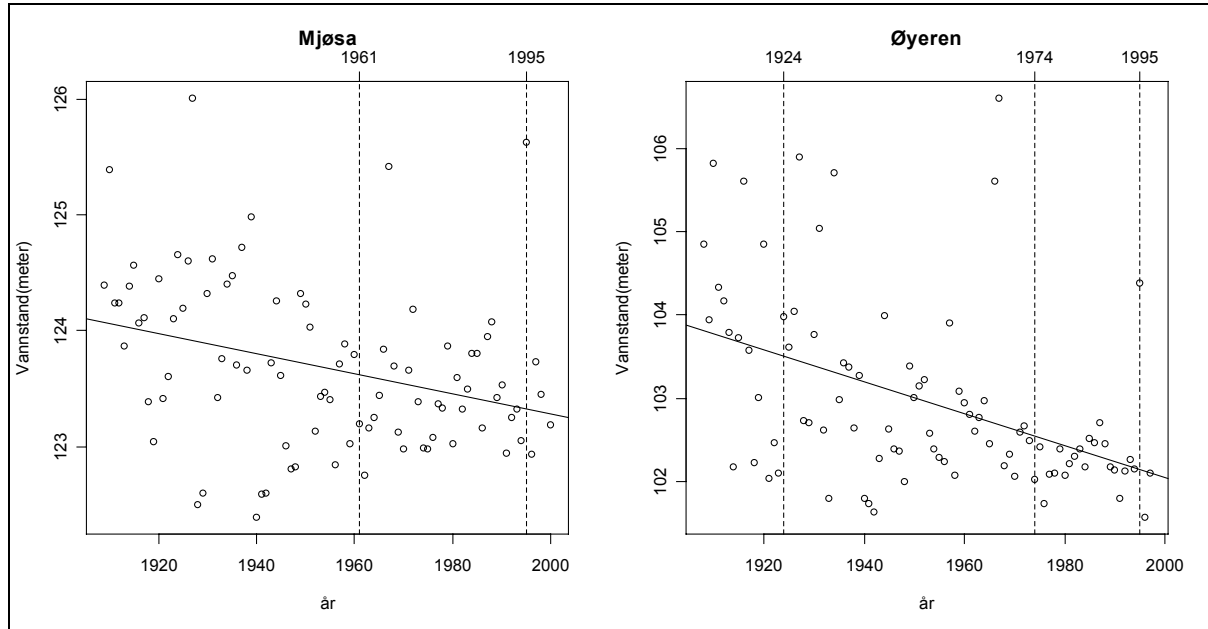


*Figur 9: Figuren over viser et autokorrelasjonsplot over vannstand fra Hamar og Lillestrøm (Mørkfoss) som sier noe om avhengigheten fra et gjennomsnittlig tidspunkt og framover i tid. 1 lag er i dette tilfellet en dag.*

Autokorrelasjonsplottet fra Figur 9 viser at dataene er sterkt avhengig av hverandre. Selv etter 45 dager er det en korrelasjon på ca. 0.5. Fra dette er det da tydelig at flere av de høyeste verdiene kommer fra samme hendelse. For POT-analysen krever dette at man bruker en kluster-metode for å luke bort de dataene som er fra samme hendelse. Bruk av POT-data er nok mer aktuelt for vannføringsdata enn i dette tilfellet hvor det er brukt vannstandsdata, fordi det er mye tregere respons i en innsjø som Mjøsa enn i for eksempel Glomma.

Det ble gjort noen forsøk på å plukke ut POT-data, men  $\lambda$ -verdiene (antall verdier per år over en bestemt terskel) var tett opptil 1, altså nesten ekvivalent med AM. Så i oppgaven er det brukt AM-verdier til den videre analysen, selv om bruk av POT er

sett på som en mer robuste metode. Det hadde vært en stor fordel å kunne brukt POT-data her siden de pålitelige måleseriene er så korte. *Petterson(1997)* har også brukt AM-data til en liknende analyse for både Øyeren og Mjøsa.



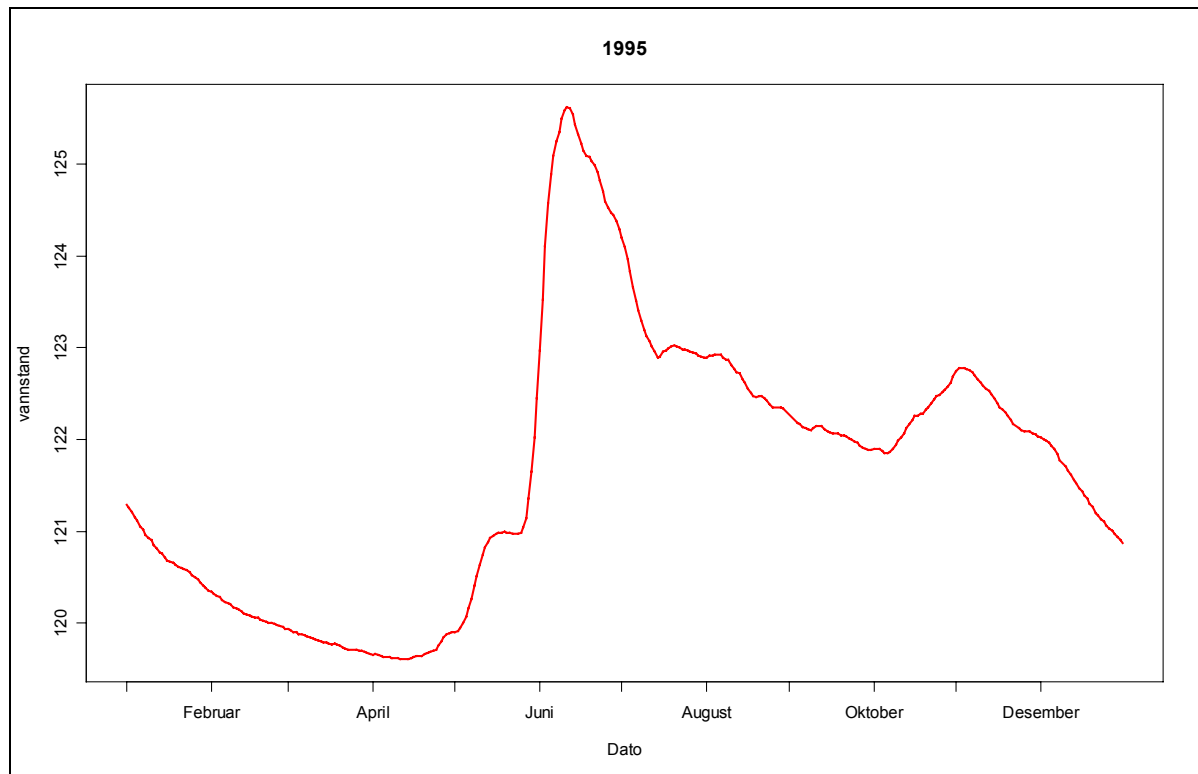
*Figur 10: Trendplott for vannstanden for målestasjonene på Mørkfoss og i Hamar*

Det er verdt å merke seg at de fleste AM verdiene kommer fra vårflommen. De største flommene i disse vassdragene kommer som regel som vårflommer. En vårflom oppstår i hovedsak pga. snøsmelting, mens en høstflom som oftest er forårsaket av regn. I Mjøsa er høstflommen som regel også et resultat av vårflommen, da det tar lang tid før vannstanden går ned etter vårflommen. Det kan være lurt å ikke blande data fra høst og vår, fordi de kommer fra to forskjellige populasjoner. Hvis det er en normal flom i vassdraget, synker vannstanden sakte. Er det derimot en storflom, går vannstanden fort ned til det nivået en normalflom ligger på.

Figur 10 viser at det er en markant negativ trend i data både for Mjøsa og Øyeren. En trend kan enten skyldes at vassdraget har blitt regulert eller at det har skjedd klimaendringer. I disse to tilfellene er det nok mest sannsynlig at trenden skyldes reguleringer.

### 5.1.1 Mjøsa

Fra Figur 10 ser vi at det er en markant trend i data, t-verdien bekrefter også dette. Siden vi er interessert i en flomfrekvensanalyse for dagens forhold er det en for stor trend i data til å bruke måleserien i sin helhet.

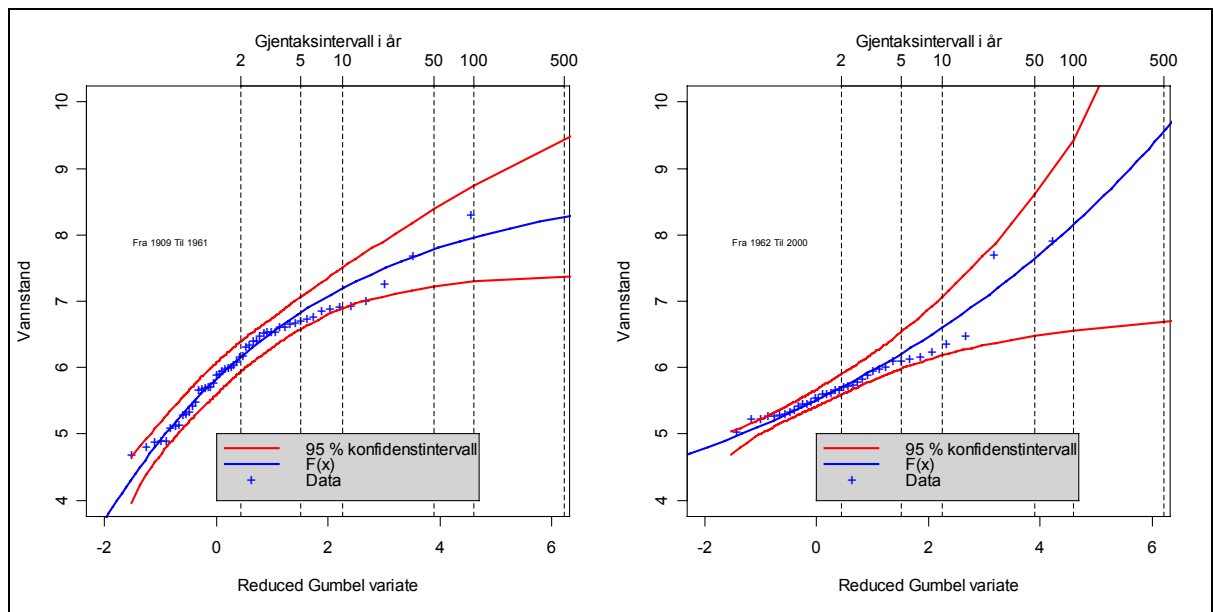


*Figur 11: Vannstandskurve for Hamar i 1995.*

Det kan se ut som at vannstanden i Mjøsa først stabiliserer seg på et fast nivå etter siste reguleringsendring i 1961. Det beste hadde da vært å bare brukt data etter 1961 i den videre analysen. Men det er ikke så lett, da den teoretisk tilpasningen ikke blir særlig robust med et så lite datagrunnlag.

Et alternativ hadde vært å fjerne trenden fra dataseriene, ved å trekke fra regresjonslinja fra Figur 10. Dette er litt skummelt da oppførselen til dataene egentlig er mer trappetrinnformet. Da måtte man eventuelt fjernet trenden for perioden frem til 1961 og koblet dette sammen med de resterende data. Det er mulig dette hadde gitt et mer reelt resultat, men man må være veldig forsiktig med å manipulere slik med data.

Det ble dermed valgt å dele opp serien i to, en periode fram til 1961 og en fra 1962 og utover, og kjøre analysen parallelt for disse to med hovedtyngden på den senere perioden.



Figur 12: Reduced gumbel-plott for Hamar for perioden 1909-61, og for perioden 1962-98.

Fra Figur 12 ser vi at plottet blir annerledes om vi plukker ut data fram til reguleringen av Svanfoss i 1961, enn etter reguleringen. Vi kan også legge merke til at den høyre kurven følger en weibullfordeling mens den venstre følger frechetfordelingen. Det kan da tyde på at det har skjedd en endring i oppførselen til vassdraget, og man bør dermed være forsiktig med å bruke data fra før 1962.

For å finne den teoretiske fordelingsfunksjonen til data ble GEV-fordelingen brukt, da teorien i kapittel 3.2.2 indikerer at det kan være lurt å bruke denne fordelingen til maksimumsverdier. Det er like viktig med en god teoretisk bakgrunn, som at fordelingen er godt tilpasset til data.

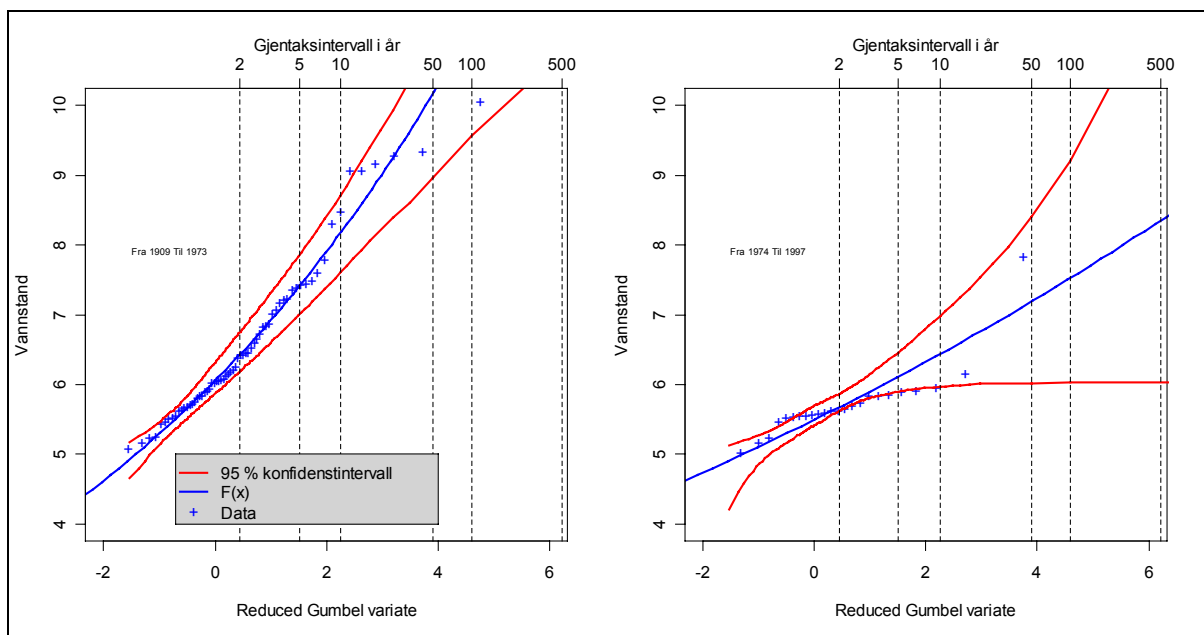
### 5.1.2 Øyeren

På samme måte som for Mjøsa ser vi av Figur 10 at det er en tydelig trend i data, faktisk større enn for Mjøsa. I motsetning til Mjøsa kan man fra Figur 13 se at reguleringer har gjort store utslag på gjentaksintervallet. Det er også verdt å merke seg at i perioden fra 1909-1973 har det vært hele 8 flommer som har nådd høyere enn hva flommen gjorde i 1995. For denne perioden er den teoretiske fordelingen godt tilpasset data, mens for den seinere perioden passer det ikke fullt så bra. Flommen i 1995 er nok den største årsaken til denne dårlige tilpasningen i tillegg til at 24 år er et lite datagrunnlag. En bedre tilpasning kan man oppnå ved å anta at 1995 er en såkalt outlier, og fjerne denne. I så tilfelle vil trolig gjentaksintervallet bli underestimert, siden det da ikke bygger på noen ekstreme hendelser. *Petterson(2002)* har brukt perioden 1903-2001 til å finne lave gjentaksintervall, og perioden 1962-2001 for å finne høye gjentaksintervall for Øyeren. En flomfrekvensanalyse for perioden 1962-

2001 gir mye større flom for høye gjentakintervall enn perioden 1974-97, selv om det meste av reguleringen i vassdraget skal være gjort innen da. Siden det er først etter 1973 at det ikke er en negativ trend i data, er det i dette arbeidet valgt å bruke perioden 1974-97 istedenfor 1962-97. Det er i tillegg valgt å se på perioden 1909-73.

Da det er nesten 3 meters forskjell på et 50 års gjentakintervall i de to analysene fra Figur 13, kan dette tyde på at man bør være meget forsiktig med hva man gjør. Selv om dataene fra 1974-97 er dårlig tilpasset GEV-fordelingen, er det nok det beste utgangspunktet for en videre flomrisikoanalyse, siden vassdraget trolig vil oppføre seg lignende i framtiden.

Det er også mulig å foreta en såkalt goodness of fit test for å verifisere om tilpasningen er god nok. En slik test vil rimelig sikkert konkludere med at tilpasningen for perioden 1974-97 ikke er god nok. Et eksempel på en slik test som fungerer greit for ekstremverdier er Anderson-Darling test. Det er også mulig å se på kjernetetthetsplott for simulering av parametere, noe som er nærmere beskrevet i kapittel 3.1.3.



Figur 13: Plottet viser flomfrekvensanalysen for Mørkfoss, delt opp før og etter siste regulering.

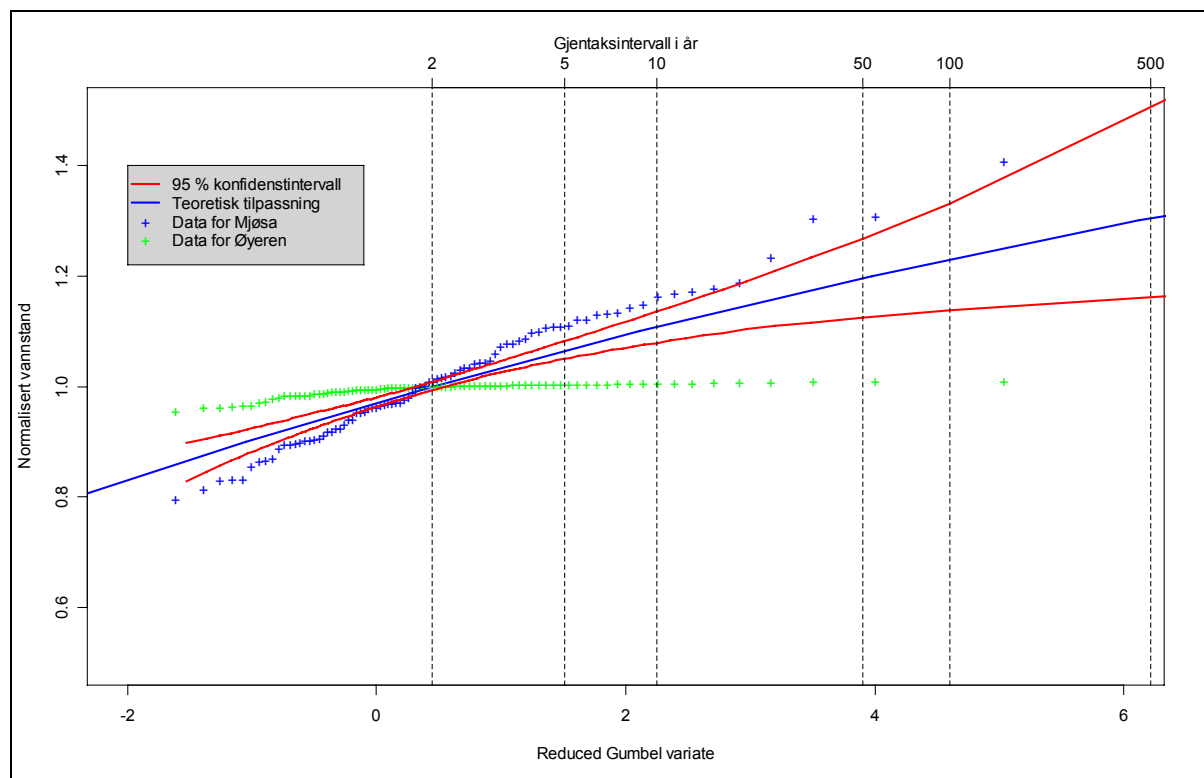
Det er også videre i risikoanalysen tatt med data fra før og etter 1974 for å sammenligne, selv om data etter siste regulering trolig er mer å stole på.

Det hadde også for Øyeren vært en mulighet å fjernet trenden vist i Figur 10, og basert den videre analysen på dette. For Øyeren hadde dette vært mer aktuelt å gjøre enn for Hamar, da perioden er kortere og det er usikkert hvor mye man kan stole på beregningen fra flomfrekvensanalysen for 1974-1997.

### 5.1.3 Regional flomfrekvensanalyse

Den opprinnelige planen for dette arbeidet var å kjøre en regional flomfrekvensanalyse for data fra Mjøsa og Øyeren, og deretter å skape skadekurver ut fra dette. Øyeren blir påvirket av både Mjøsa- og Glomma-vassdraget, slik at forholdene i Øyeren blir så forskjellig fra Mjøsa at en regional flomfrekvensanalyse blir vanskelig. Under store flommer er det vanlig at man holder igjen vannet i Mjøsa, slik at flomtoppen fra Vorma og Glomma ikke skal møtes samtidig og skape problemer i Øyeren

Det ble gjort noen forsøk med en slik regional analyse, men resultatet viste seg å bli for sprikende. En eventuell fordel med en regional analyse hadde vært at man hadde fått et større datagrunnlag, og kunne minimert usikkerheten.



Figur 14: Et reduced Gumbel plott over regional flomfrekvensanalyse for Mjøsa og Øyeren for perioden 1909-1994

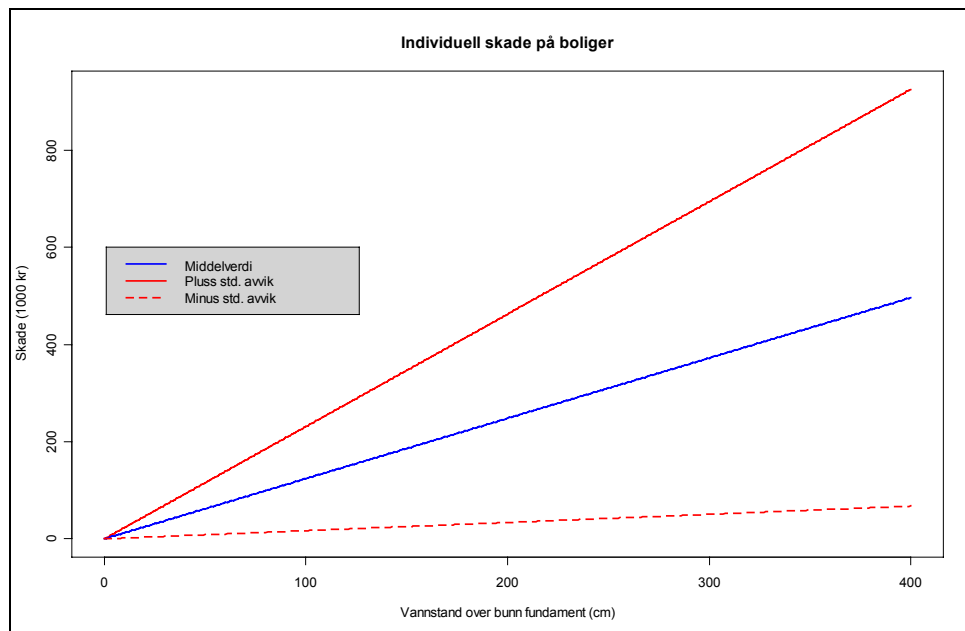
Figur 14 viser at den regionale tilpassingen, som også er med AM-data, ikke er helt bra og seriene spriker litt for mye fra hverandre. Det er tydelig å se at de ikke følger den samme fordelingen. I en regional flomfrekvensanalyse er det også en fordel å bruke så mange målestasjoner som mulig, to stasjoner kan være litt lite hvis data ikke passer godt overens.



## 5.2 Usikkerhetsberegninger

### 5.2.1 Usikkerhet fra kostnadberegninger

Det er valgt å bruke data fra en regresjonsanalyse utført etter flommen i 1995 til beregningen av den individuelle skaden til hus ( $l'(y)$ ), som er beskrevet i likning 1. Fra denne regresjonsanalysen er det litt lite data tilgjengelig til å si noe om usikkerheten, men det er allikevel valgt å legge med Figur 27 hvor det er implementert usikkerheten fra den lineære regresjonsanalysen til *Wathne m. flere (1999)*. De har valgt å kun utføre en enkel lineær regresjonsanalyse. Hadde man hatt tilgang til disse dataene kunne man kanskje også forsøkt å tilpasse ikke-lineært, i hvert fall hvis det hadde vært mer data tilgjengelig. I tillegg kunne man også her beregnet usikkerheten ut i fra bootstrapping.



Figur 15: Figuren viser regresjonsanalysen gjort av *Wathne m. flere (1999)*, for skade på individuelle hus.

Figur 15 viser at det er en stor usikkerhet med en slik regresjonsanalyse. Den nedre usikkerheten har lave kostnadsverdier, og den vil gjøre store utslag på konfidensintervallet til skadekurven. Standardavviket er 1.24.

Siden usikkerheten i beregningene fra Figur 15 er store foreslår *Gottschalk og Kraskovskaia (1999)* at man istedenfor å bruke hele regresjonsanalysen til å finne  $l'(y)$ , deler opp i to trinn, høy og lav skade:

Lav skade :  $-2.5 < y < -0.5$

Høy skade :  $-0.5 < y$

(61)

Forskjellen på resultatene fra disse to beregningsmåtene for individuell skade vil bli nærmere studert i kapittel 5.3.2

I dette arbeidet vil likning 1 bli brukt som standard i risikoanalysen for bolighus.

Skadetallene som er samlet inn kommer fra elva Glomma, hvor det som kjent er rennende vann. I for eksempel Hamar vil vannet holde seg mer i ro, og da vil trolig skaden bli mindre i forhold til et tilsvarende hus ved Glomma. Dette er ikke verifisert.

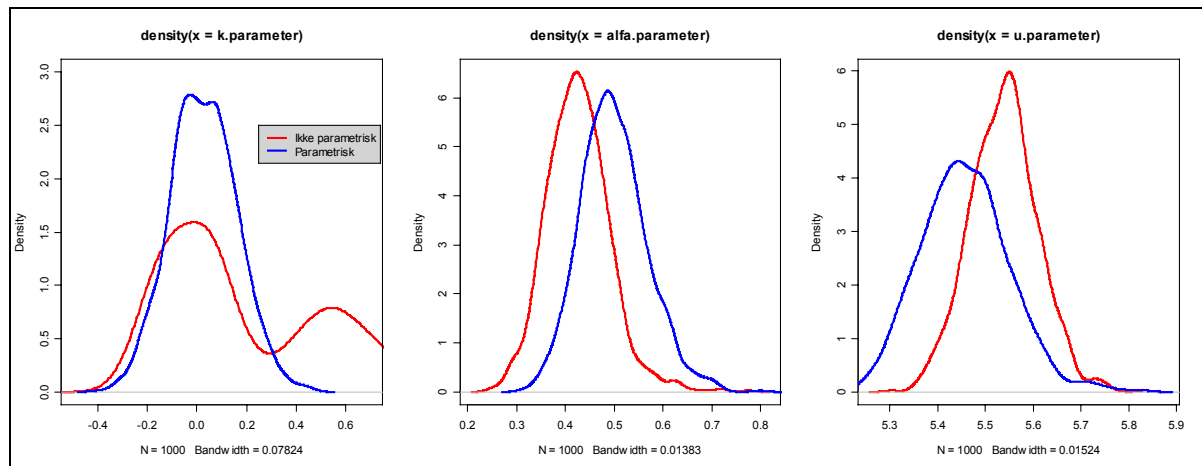
### **5.2.2 Usikkerheten fra flomfrekvensanalysen**

Ut fra det ovenstående er det da lettere å si noe om usikkerheten på bakgrunn av flomfrekvensanalysen. Beregning av usikkerheten til parameterne fra den teoretiske tilpasningen til data, er et utbredt tema i statistikken.

I dette arbeidet er man i tillegg interessert i usikkerheten til fordelingsfunksjonen i seg selv, da dette er eneste måten å koble usikkerheten til skadekurven på.

Til usikkerhetsanalysen er det brukt bootstrapping som er beskrevet i kapittel 3.3.1.

Det ble valgt å gjøre et forsøk på å bruke simulering i bootstrappanalysen. For at man skal kunne bruke simulering kreves det at et kjernetetthetsestimeringplott for parameterne funnet ut fra parametrisk bootstrapping er rimelig likt et tilsvarende plott for ikke-parametrisk bootstrapping. Diagrammene i Figur 16 tilsier dermed at GEV-tilpasningen ikke er god nok og man burde bruke ikke-parametrisk bootstrapping (resampling). Dette betyr egentlig også at tilpasningen av data til GEV-fordelingen ikke er god nok, men som tidligere nevnt i kapittel 5.1 så er denne tilpasningen trolig det beste alternativet man har til en flomfrekvensanalyse for Lillestrøm. Alt beskrevet her er for perioden etter siste regulering. Hamar gir tilsvarende svar som for Lillestrøm. Konklusjonen fra dette blir at det i det videre arbeidet blir brukt resampling til bootstrappingen. En goodness of fit test kan også brukes til å verifisere om man kan bruke simuleringsbasert bootstrapping. Hadde man valgt å bruke vannstandsdata fra perioden før siste regulering, ville man kunne brukt simulering til bootstrappinganalysen, fordi tilpasningen er mye bedre her.



Figur 16: Kjernetetthetsestimeringsplott for parameterne for Øyeren, gitt data for perioden 1974-1997 for Øyeren.

	$k$	$\alpha$	$u$
Øvre konfidensintervall	0.17	0.50	5.63
<i>L</i> -moment estimering	-0.14	0.42	5.51
Nedre konfidensintervall	-0.28	0.29	5.36
Bias	0.08	0.02	0.02
Maximum likelihood estimering	-0.18	0.34	5.54

Tabell 5.2.1 Konfidensintervaller til parameterne for Mjøsa gitt data for perioden 1962-2000.

Usikkerhetsestimatene i Tabell 5.2.1 og Tabell 5.2.2 ble beregnet ut i fra ikke-parametrisk bootstrapping, og  $BC_a$ -metoden for konstruksjon av konfidensintervall.

Vi ser at forventningsskjevheten (bias) til parameterne er nesten lik null, det vil si at den gjennomsnittlige verdien til bootstrappsamplene er omtrent lik den estimerte verdien.

Det er interessant å merke seg at konfidensintervallet til formparameteren  $k$  ligger på begge sider av 0. Det vil si at dataene enten kan følge en frechet-, weibull- eller gumbelfordeling. Hvis data følger weibullfordelingen, vil da vannstanden aldri komme over et visst nivå.

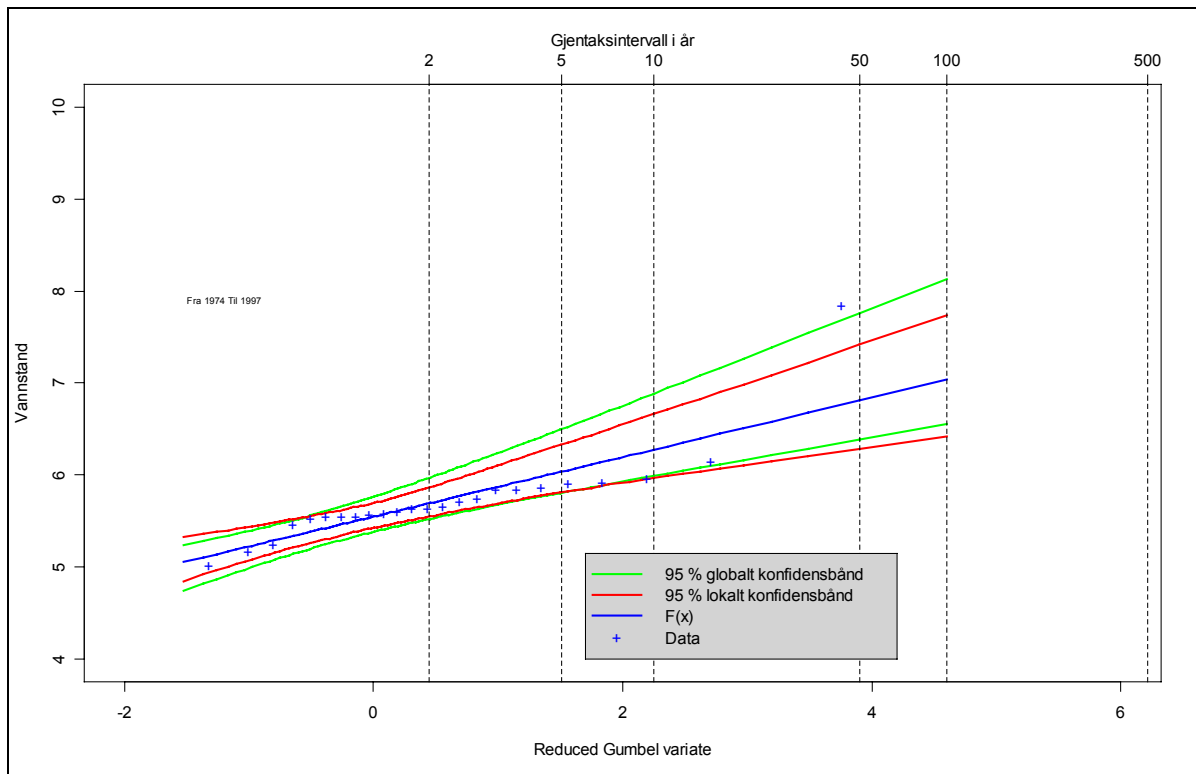
Tabell 5.2.1 og Tabell 5.2.2 viser kun en oversikt over konfidensintervallene til parameterne, og de kan ikke brukes til å si noe om usikkerheten til hele skadefunksjonen. For å si noe om denne usikkerheten må det lages et såkalt konfidensbånd til fordelingsfunksjonen. Dette er i oppgaven konstruert på bakgrunn av simulering av data 1000 ganger, som nevnt i kapittel 3.3.4.

	$k$	$\alpha$	$U$
<i>Øvre konfidensintervall</i>	0.57	0.50	5.59
<i>L-moment estimering</i>	-0.05	0.40	5.50
<i>Nedre konfidensintervall</i>	-0.32	0.28	5.32
<i>Bias</i>	0.26	0.03	0.05
<i>Maximum likelihood estimering</i>	-0.06	0.32	5.54

Tabell 5.2.2: Konfidensintervaller til parameterne for Øyeren gitt data for perioden 1974-1997.

Konfidensbåndene som er brukt i oppgaven er lokale. Hadde man derimot brukt maximum likelihood til parameterestimeringen, hadde det vært oppnåelig i tillegg å konstruere et globalt konfidensbånd. I *Cheng og Iles(1983)* står det beskrevet hvordan man kan lage et globalt konfidensbånd for weibull- og gumbelfordelingen, som trolig kan skrives om slik at det kan brukes for frechetfordelingen også.

Selv om maximum likelihood er sett på som en ikke fullt så god estimeringsmetode for ekstremverdifordelinger som L-momenter, så bør dette ikke absolutt bety at man ikke skal bruke denne metoden til dette formålet. Maximum likelihood har flere fordeler kontra L-momenter. Den er en eldre metode enn L-momenter som kun har eksistert i ca. 10 år. Likelihood-metoden er også utbredt innenfor statistikken, og det er utarbeidet mange metoder som bygger på maximum likelihood estimering. Etter hvert kommer trolig også L-momenter å bli mer utbredt, og statistikerne vil utarbeide metoder som gjør bruk av L-momenter mer attraktivt.



Figur 17: Flomfrekvenskurve der vannføringsdata for Mørkfoss er tilpasset gumbelfordelingen. Det er implementert både et globalt og et lokalt konfidensbånd. Det er brukt maximum likelihood som estimeringsmetode.

Under maximum likelihood estimering for data for Mjøsa er  $k$ -verdien i GEV-fordelingen tett opptil 0, slik at data nesten følger en gumbelfordeling. Det ble dermed gjort noen forsøk på å konstruere et globalt konfidensbånd ved å tilpasse data til maximum likelihood estimering for gumbelfordelingen. Dette er vist i Figur 17. Det globale båndet er litt bredere enn det lokale, noe det også skal være i følge teorien.

En tilpasning til GEV-fordelingen beskriver en mye større usikkerhet enn å tilpasse til gumbelfordelingen siden denne har tre parametere mot to for gumbel.

### 5.2.3 Usikkerhet fra GIS-analysen

Av de mange usikkerhetsfaktorene er det greit å nevne at GIS-analysen også bringer med seg en god del usikkerheter. Alle risikoberegninger i oppgaven er beregnet med bakgrunn i høydeverdier interpolert ut i fra høydekurver. Det kan ofte være litt risikabelt å interpolere fra høydekoter. Hvis det blir brukt gal interpoleringsmetode kan dette gi en trappetrinnformet høydemodell. Det ble i dette arbeidet brukt en interpolasjonsmetode kalt topogrid, som er beregnet på bruk innen hydrologi. I tillegg

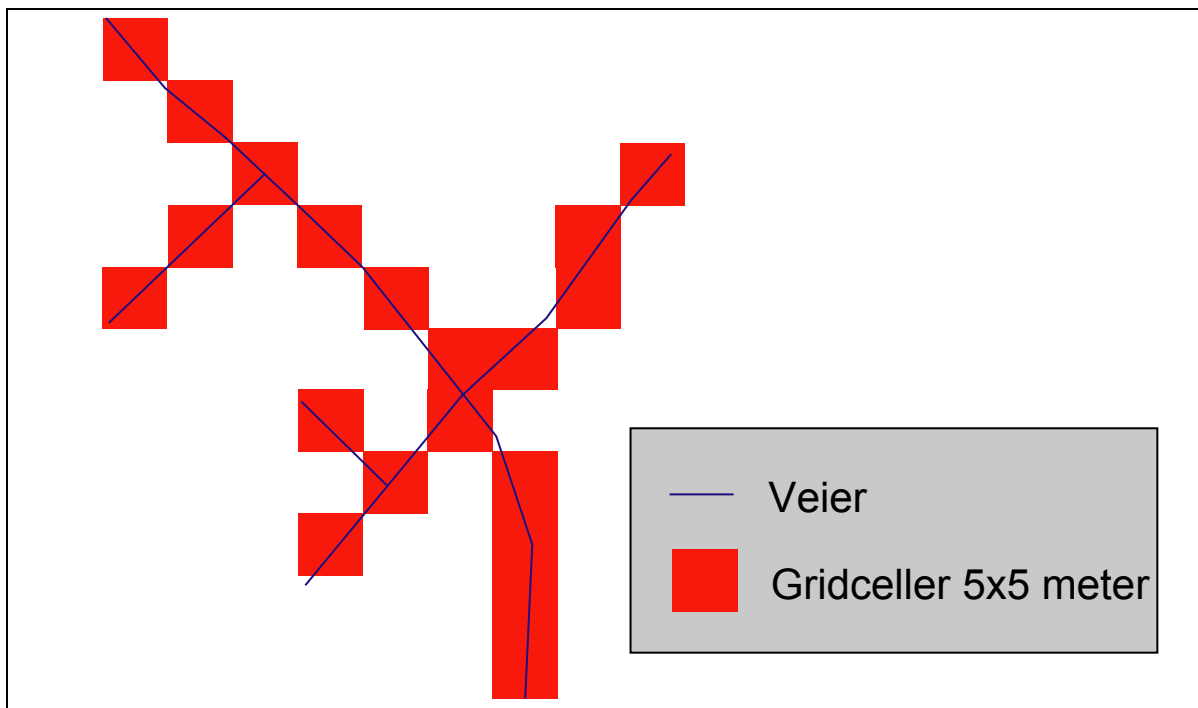
er den integrert i Arcinfo. Det er ikke gjort noen test på denne metoden, bare antatt at den er grei. Denne metoden er ytterligere beskrevet i *Sælthun m. flere (1999)*.

Det beste hadde nok vært å brukt en terrengmodell direkte basert på flyfoto eller laser/rader, og ikke en som først er konvertert til høydekurver og så videre til en terrengmodell. I en risikoanalyse er det ekstremt viktig at kartgrunnlaget er godt. Plasseringen av de forskjellige flomutsatte objektene er mer enn god nok, men det er høydeverdiene deres som er vanskelig å beregne eksakt.

Høydeverdiene blir nok best tilpasset for Hamar, da dets kartgrunnlag har et 1 meters kotegrunnlag mot 5 meter i Lillestrøm. Standardavviket til høydeverdiene kan antas å være 30% av ekvidistansen(*Gottschalk og Krasovskaia (1999)*).

For å kunne beregne skade på jernbane og veier, ble dataene som var presentert som linjer i sosi-filene tatt inn i GIS-verktøyet Arcinfo og gjort om til 5-meters gridceller. Dette for å ha mulighet til å tilordne høydeverdier. I tillegg blir det mye greiere å regne med når alle elementer er like lange. Når man har dette er det bare å telle opp antall gridceller som er oversvømt og gange med gridstørrelsen, og man vet hvor mange kilometer som er oversvømt. Dette er ikke den mest nøyaktige metoden, men skal det gjøres mer nøyaktig er det fort gjort å støte på problemer.

Det ble før det ble bestemt å dele opp i gridceller gjort forsøk på å dele opp linjene i like stor biter, men da var det ikke mulig å tilordne høydeverdier. Arcinfo er ikke et 3D-dimensjonalt GIS-verktøy, noe som antakelig hadde vært til nytte for å få dette til.



Figur 18: Figuren viser at gridstørrelsen ikke samsvarer helt med den reelle lengden på veien.

Konverteringen til gridceller medfører da at vei/jernbane vil bli presentert i trappeform istedenfor rette linjer, slik Figur 18 viser. En vei som da går for eksempel diagonalt igjennom en gridcelle blir da registrert som om den er 5 meter selv om den i virkeligheten enten er lenger eller kortere. Det er mulig dette kan justeres med en faktor. Et annet problem kan oppstå for eksempel ved en jernbanestasjon der gridcellene kan smelte sammen slik at to skinner kun blir registrert som en. Dette vil da kanskje kunne unngås med å forandre gridcellestørrelsen til for eksempel 1 x 1 meter, men forsøk på dette viste ingen tydelige forbedringer. I tillegg kreves det veldig mye regnekraft for å lage en gridmodell som er 1 x 1 meter.

Høydeverdiene til jernbane/vei er kun bestemt ut i fra en høydemodell, slik at der jernbane/vei går i bru over vann, vil høydeverdien til jernbanen/veien bli den samme som vannet, slik at det her vil bli registrert skade selv om det ikke er noe. Dette kan enklest unngås ved å i ettertid trekke i fra den kostnaden som avstanden til vei/jernbane som går i bruer gir.

#### 5.2.4 Generelt om usikkerhetsberegningene

Siden dette arbeidet er konsentrert rundt L-momenter, har dette medført at det kun er implementert lokale konfidensbånd til skadekurvene. Årsaken er i tillegg at det er greit å implementere og fordi noe annet fort blir for komplisert. Disse konfidensbåndene er blant annet vist i Figur 12 og Figur 13.

Slike usikkerhetsberegninger tar fort veldig langt tid å beregne i et dataprogram, slik at det i oppgaven kun er brukt 1000 simuleringer selv om 10 000 hadde vært optimalt. Det er konstruksjonen av usikkerheten til skadekurven som gjør at det tar så lang tid. I tillegg kan høydeverdiene som brukes til å skape skadeplottene, bli delt opp i 10 slik at det ikke skal ta for lang tid å beregne. Dette medfører da dessverre at oppløsningen på plottene ikke blir helt gode. Alle plottene gjengitt i oppgaven er plottet med oppløsning på 100.

## 5.3 Risikoanalyse

### 5.3.1 Lillestrøm

Det er blitt gjort mye arbeid for å sikre mot flom i sentrumsområdet av Lillestrøm. Hele bykjernen skal være sikret mot en 300 års flom i følge NVEs beregninger. Altså blir ikke Lillestrøm rammet hvis vannet ikke stiger høyere enn 106 m. o. h. Storofsen i 1789 ga en vannstand på 111.538 m. o. h, noe som i dag hadde satt hele bykjernen under vann, og 10.000 mennesker hadde blitt berørt. Det var i 1798 uten tiltak. Dette

ville i dag gitt lavere verdi. Man har antatt at det blir minimalt med skade på Lillestrøm hvis vannstanden er under 106 m. o. h, men med en gang den kommer over 106 m. o. h blir det enorme skader.

Kartdataene for Lillestrøm var ikke så gode som de for Hamar, de manglet mye nyttig informasjon. Det manglet blant annet polygondata for bolighus, slik at det ikke var noe informasjon tilgjengelig om hva slags bygning det var. Det er dermed kun gjort flomrisikoanalyse på jernbaner og veier i område.

Lillestrøm er et viktig knutepunkt når det gjelder jernbanetrafikk i Norge, da all trafikk nordover og til Stockholm går igjennom dette område. En eventuell storflom vill skape problemer for mye av jernbanetrafikken. For veitrafikken vil en flom omtrent bare skape lokale problemer. Det er ingen store, trafikkerte veier som går gjennom området.

### *Situasjon for mulig flom i Lillestrøm*

For at man i Lillestrøm skal overstige flomvollene rundt byen som er på 106 meter, krever dette :

1. Veldig sen og stor vårflom fra skogsområdene rundt Oslo, altså mye vann i Nittelva og Leira
2. Spesielle omstendigheter som gjør at det ikke er mulig å holde igjen vann i Mjøsa
3. Som konsekvens av pkt. 2 kommer flomtoppen i Vorma og Glomma samtidig
4. Like mye vann som under flommen i 1995.

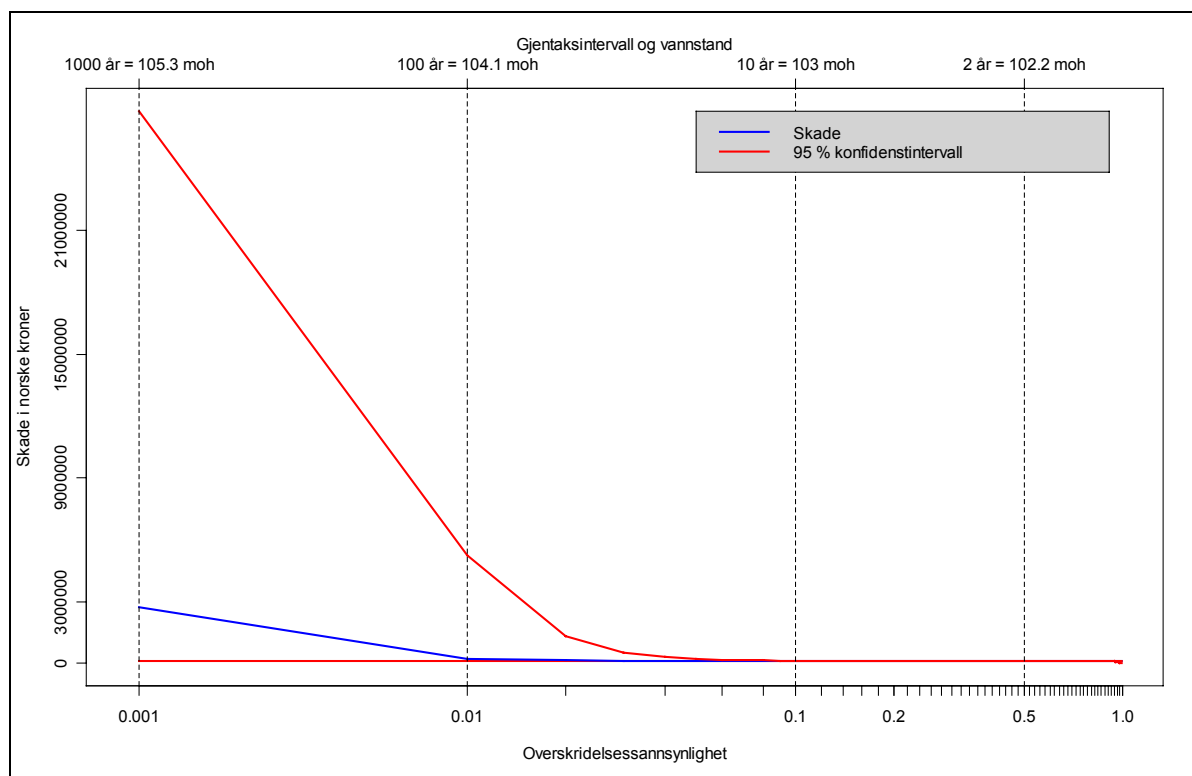
Eventuelt:

5. Brist i flomvollen.

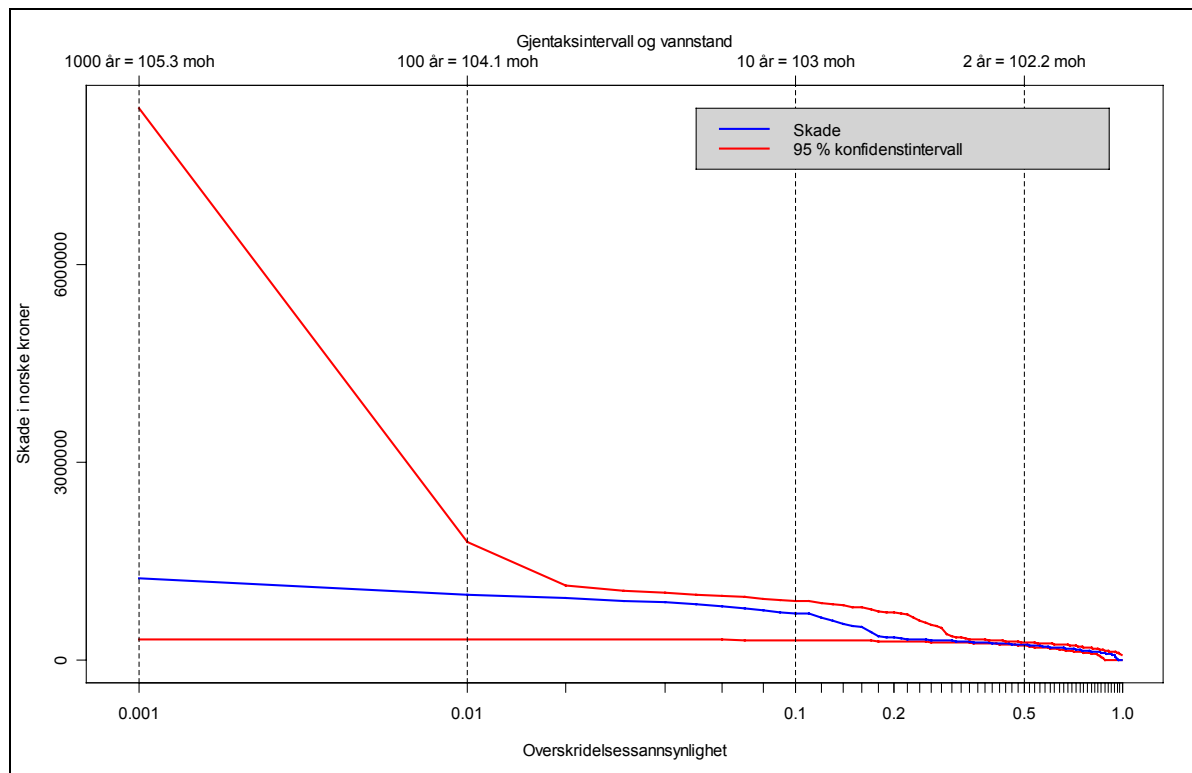
Etter storflommen i 1927 ble det satt opp en god del flomvoller rundt byen, men før flommene i 66 og 67 var flere av disse fjernet fordi folk trengte denne plassen. Dette er ting som fort kan skje igjen. Hvis det ikke blir flom på noen år kan pumpestasjonen også bli dårlig vedlikeholdt, og en flom kan fort overraske beboerne. Som beskrevet i kapittel 5.1.2 så er sannsynligheten mindre for å få flom nå enn før, slik at det kan gå flere hundre år før man får en stor flom igjen. Innen dette inntreffer, kan det ha blitt gjort mange forandringer og det kan få katastrofale følger, for en dag kommer storflommen igjen.

Selv om det over antydes at det er liten sannsynlighet for at det skal bli flom i Lillestrøm igjen, så er det i dette arbeidet konstruert et par skadekurver for vei og jernbane. De fleste kurvene antar da at det ikke er noen flomvoll rundt byen.



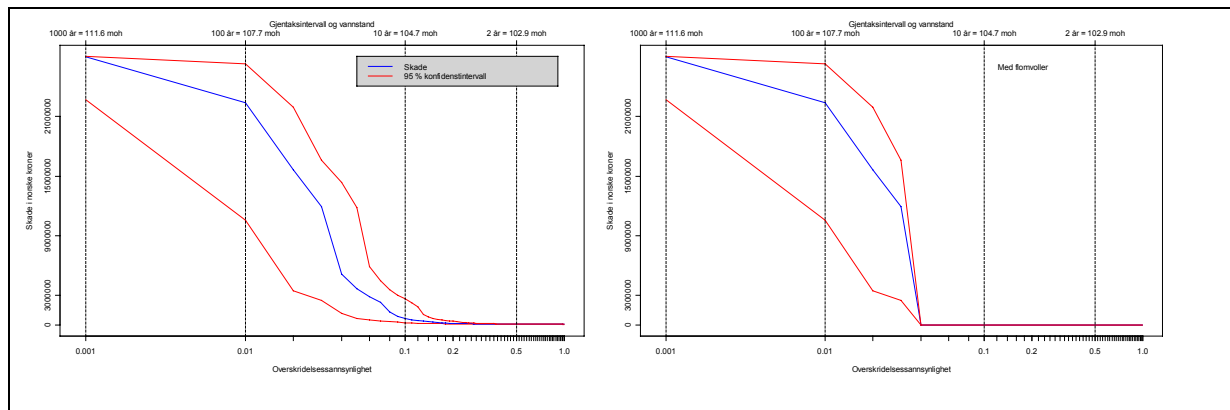


Figur 19: Skadekurve for veier i Lillestrøm, antatt at det ikke er hjelp i flomvollene rundt byen. Kurven bygger på en flomfrekvensanalyse med data fra 1974-1997.



Figur 20: Skadekurve for jernbane i Lillestrøm, antatt at det ikke er hjelp i flomvollene rundt byen. Kurven bygger på en flomfrekvensanalyse med data fra 1974-1997.

Kurvene som er vist i Figur 19 og Figur 20 er konstruert på bakgrunn av at det ikke er flomvoller til stede. Skulle man tatt med dette i analysen måtte man implementert vannstandsdata som i følge data fra 1974-1997 er høyere enn for en 1000 års flom, noe som hadde gitt alt for høy usikkerhet. Det kan da sies at det kreves en større flom enn en 1000 års flom for å oversvømme Lillestrøm. Man må derfor se bort fra flomsikring, noe som ikke er bra for å få reelle tall. Figur 21 viser at en analyse med data fra 1909-1973 er så forskjellig fra Figur 19 at det er mulig å lage en kurve hvor flomvollene er implementert.



Figur 21: Skadekurve for veier i Lillestrøm som bygger på en flomfrekvensanalyse for perioden 1909-1973. I Kurven til høyre er det tatt hensyn til flomvollene rundt byen.

Det er også verdt å merke seg at skaden gjengitt i Figur 21 blir mye større enn skaden fra Figur 19. Dette er da et tydelig bevis på at reguleringsarbeid både oppstrøm og nedstrøms har gitt resultater, og byen er på langt nær så utsatt nå som før. Det viser også tydelig at det er skjedd noe med frekvenskurven. Problemet for risikoanalysen blir da om man skal stole fullt ut på analysen beregnet ut fra perioden 1974-97. Det kan nok være greit å ha i bakhodet at dette **kan** være underestimert for ekstreme flommer, siden det er få store flommer i denne perioden.

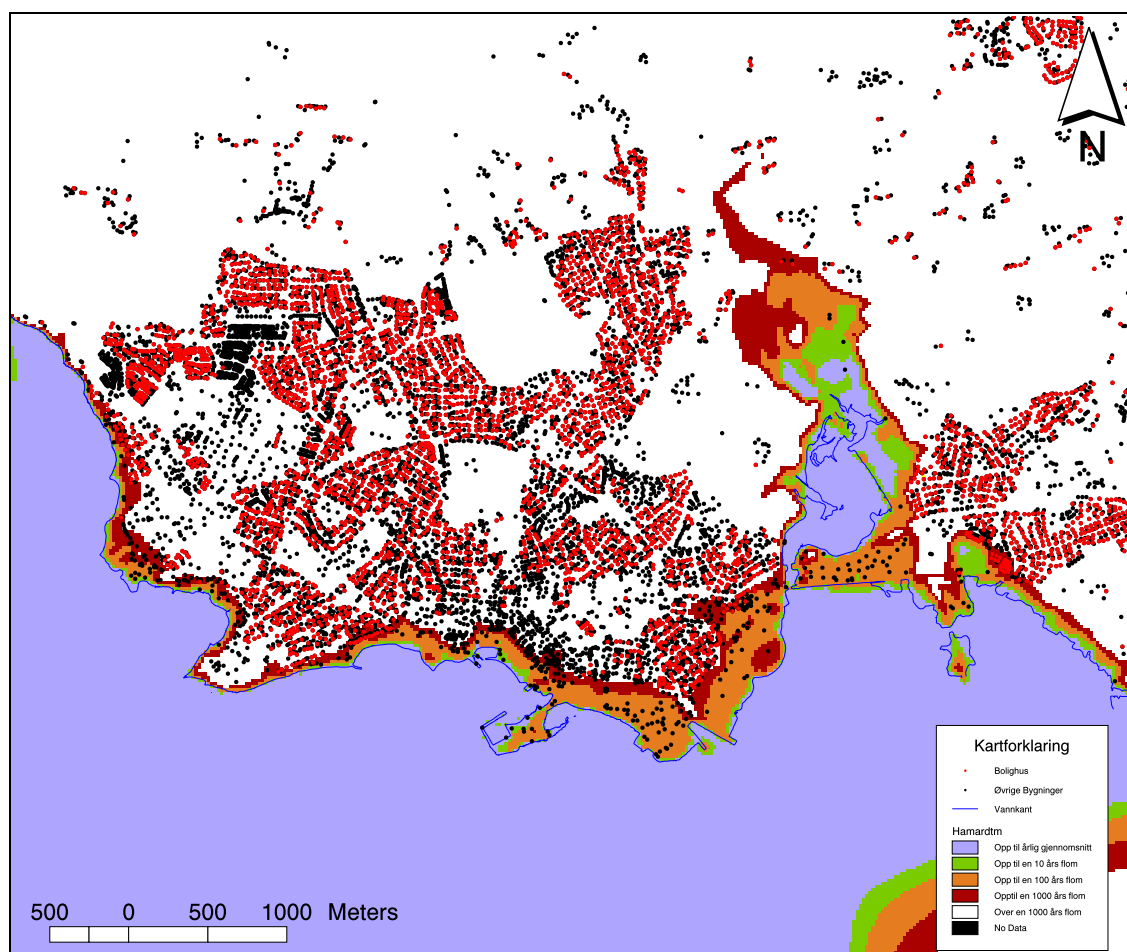
Tallene blir med en gang mer troverdig hvis vannet skulle komme inn i bykjernen. Den forventede skaden for veier i Lillestrømområdet blir for alle tilfellene beskrevet over, lik 0 norske kroner. Den forventede skaden for jernbane blir 676 000 NOK, beregnet med utgangspunkt i vannstandsdata fra perioden 1973-1997. Dette tallet er nok litt lavere, fordi der jernbanen går i bru så blir høyden på banen satt til gjennomsnittlig vannhøyde fordi bruer ikke er implementert i terrengmodellen.

### 5.3.2 Hamar

Hamar har ikke vært så utsatt for flom som det Lillestrøm har vært. Hamar ligger mer i en helling ned mot vannet, mens Lillestrøm ligger på en elveslette med elver på alle kanter. I Hamar er det også områder som er utsatt, men disse er først og fremst industriområder. Det er dermed gjort mye mindre for å sikre mot flom i Hamarområdet, det er kun satt opp noen pumpestasjoner. I tillegg A er det presentert en oversikt over hvor mange av dagens hus som ville blitt oversvømt hvis

vannstanden hadde kommet like høyt som det gjorde i 1789 og 1995. Tabellen er beregnet ut i fra tilgjengelig SOSI-data og vannstandsverdier fra Vesle- og Storofsen.

Innen risikoanalysen var det mest informasjon tilgjengelig for Hamar. Det ble regnet ut risiko for bolighus i tillegg til vei og jernbane. Fra Figur 22 ser man at de fleste bolighusene er plassert et godt stykke unna Mjøsa. Det er tydelig å se at det er flere andre bygningstyper som er mer utsatt, og som ville gitt mer interessante skadetall. Fra kapittel 2.3 er det verdt å merke seg at skadefunksjonen er beregnet slik at et hus vil bli påført skade når vannstanden ligger 2.5 meter lavere enn selve huset, slik at flere hus vil bli påført skade på kjellere selv om ikke vannet når opp til selve huset.

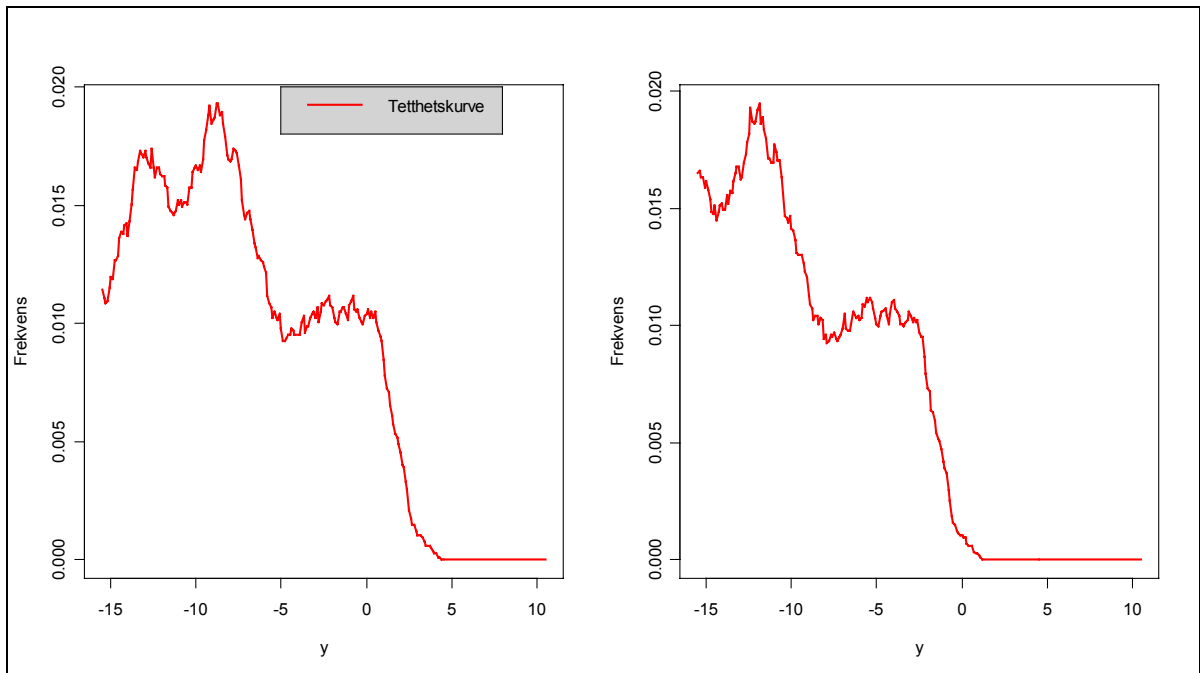


Figur 22: Figuren viser husfordeling for Hamar sentrum. De røde punktene er eneboliger, mens de svarte er øvrige bygningstyper. Bygninger er ikke presentert i sin reelle størrelse her.

### Skadefunksjonen for bolighus

For å lage en skadekurve for bolighus er det brukt metoden beskrevet i kapittel 4.4. Dataene presentert i Figur 23 legger grunnlaget for skadekurven og er et plott på tetthetsfunksjonen presentert i likning 50. Den kritiske vannstanden,  $y = z-h$  er positiv

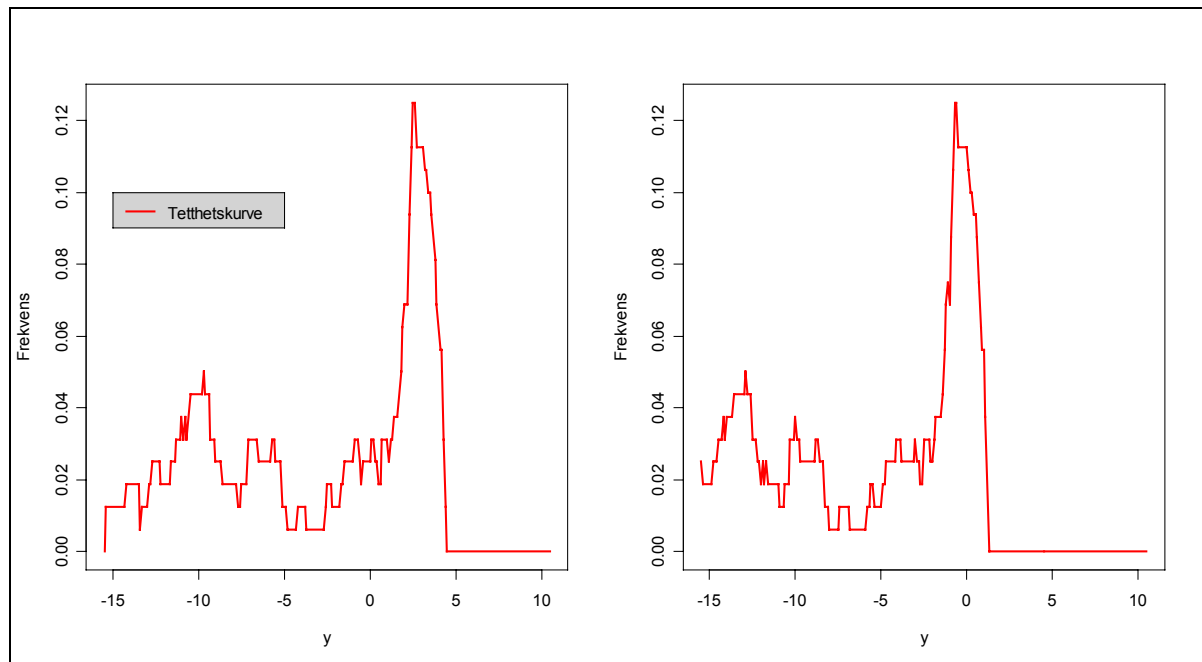
når et hus er rammet av flom og negativt når det ikke er det. Et hus er antatt flomutsatt når vannet er 2.5 meter under bakken og oppover. Det er tydelig å se av de to plottene i Figur 23 at andelen flomberørte hus ikke øker betraktelig selv om det er en forskjell i vannstand på nesten 5 meter. Dette bekrefter også Figur 22 som sier at få av husene i dette området er flomutsatte bolighus.



Figur 23: Fordelingsfunksjon for kritisk lokal vannstand,  $f_Y(y_c | z)$  for bolighus (111). For den høyre kurven er det gitt en vannstand på 122.69, mens det er gitt en vannstand på 125.86 for kurven til venstre.

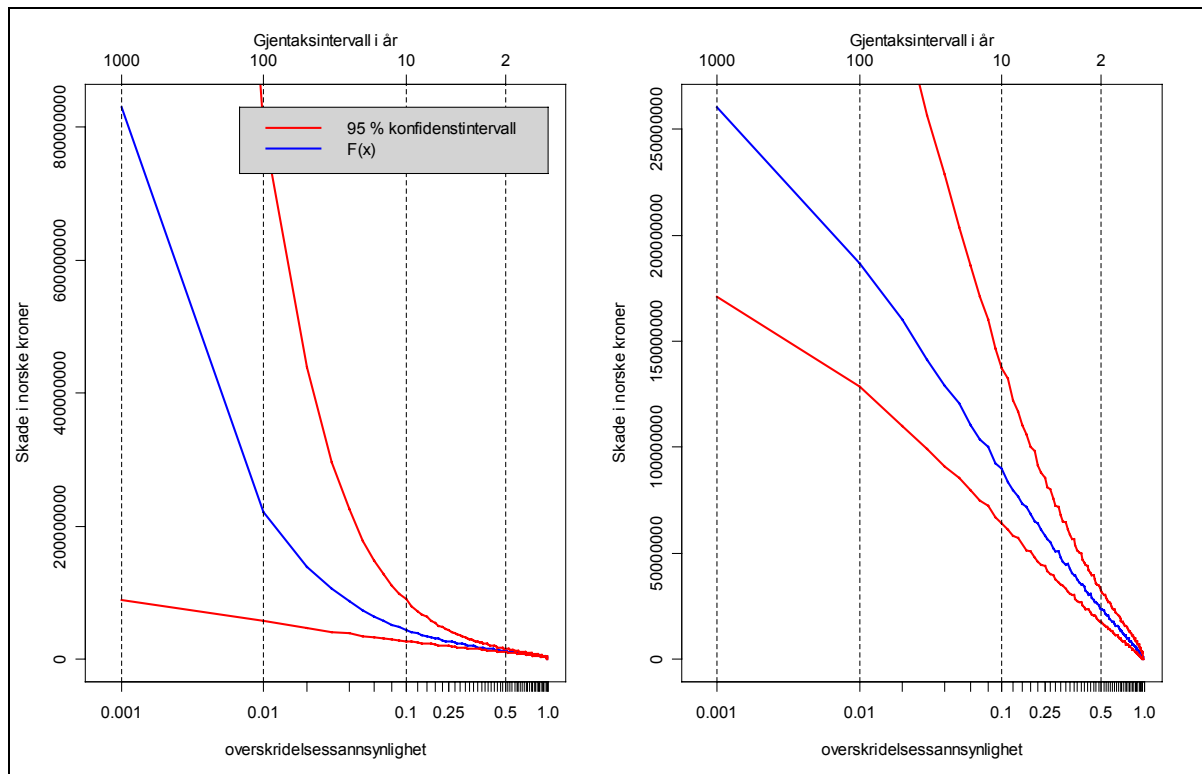
Figur 22 gir en mer interessant kurve enn hva figur 21 gir. Her er det verkstedbygninger som er plottet, og det er tydelig å se at denne type bygninger er mer flomutsatt. Det er dessverre ingen skadefunksjon tilgjengelig for denne typen. Fra figur 22 kommer det fram at verkstedsbygningene i område er konsentrert rundt havna, og omtrent hele populasjonen blir oversvømt ved høy vannstand. Fra figur 21 er det tydelig å se at antall hus blir større jo lenger man kommer unna. Dette vil for høyere verdier av  $y$  gå ned igjen, fordi elementær statistikk sier at arealet under en frekvenskurve alltid skal være 1.

Det er mulig at figur 22 er litt misvisende siden det trolig ikke er så vanlig at slike bygg har kjeller. I beregningene er  $y = 0$  der kjelleren starter og den er antatt å være 2.5 meter høy. Det er ikke foretatt noen undersøkelser på kjeller i slike lagerbygg, så derfor er det brukt samme kjellerantakelse i denne figuren som for bolighus.



Figur 24: Fordelingsfunksjon for kritisk lokal vannstand,  $f_Y(y_c | z)$  for verkstedsbygninger (212). For den høyre kurven er det en gitt vannstand på 122.69, mens det er gitt en vannstand på 125.86 for kurven til venstre.

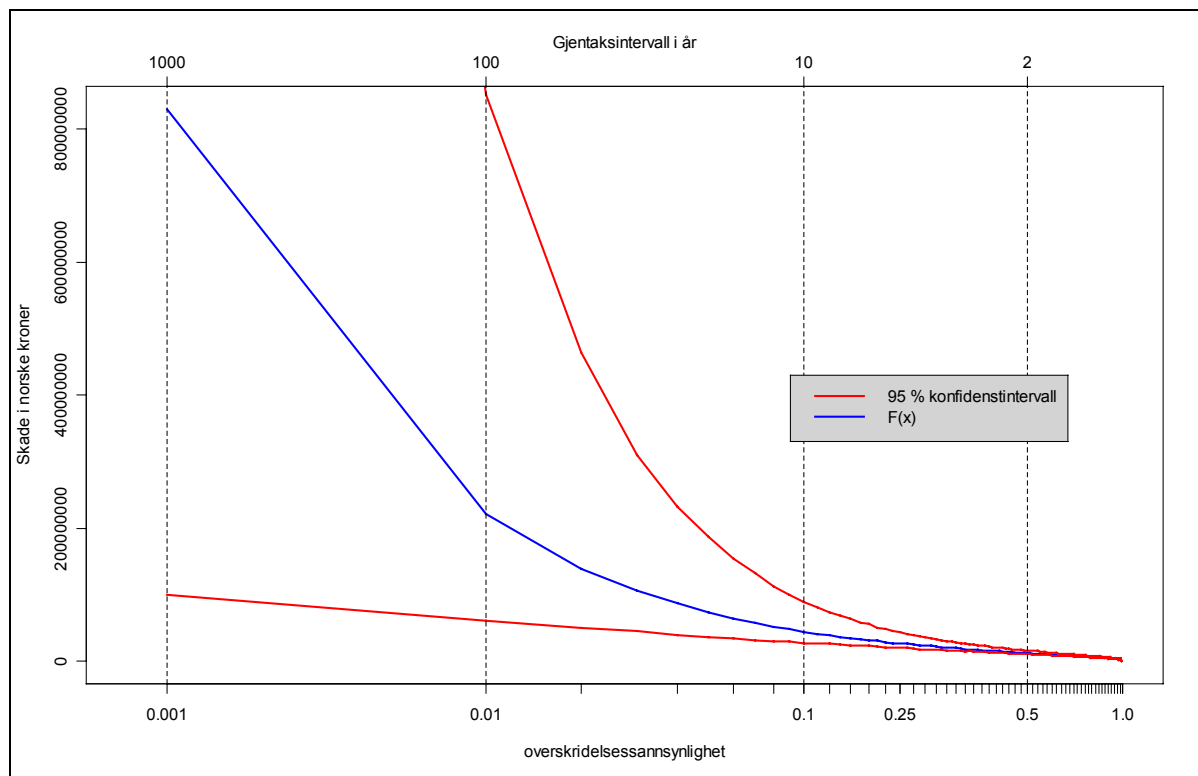
Det er i oppgaven regnet ut skadefunksjon både for perioden fram til 1961, og fra 1962 til 2000. Dette er presentert i Figur 25, og det er tydelig å se at resultatet blir vesentlig forskjellig for de to periodene. Konfidensbåndet til den seneste perioden er mye bredere enn det andre, noe som samsvarer med Figur 12. En interessant observasjon er at perioden etter siste regulering er den som gir de høyeste skadetallene, dette skyldes at Figur 12 antyder at den andre perioden konvergerer mot en gitt vannstand for høye gjentakstintervall, altså den følger en weibullfordeling. Som beskrevet i kapittel 5.1 vil frekvensanalysen for data etter siste regulering være mest interessant, og vil i de påfølgende plottene bli brukt som grunnlag.



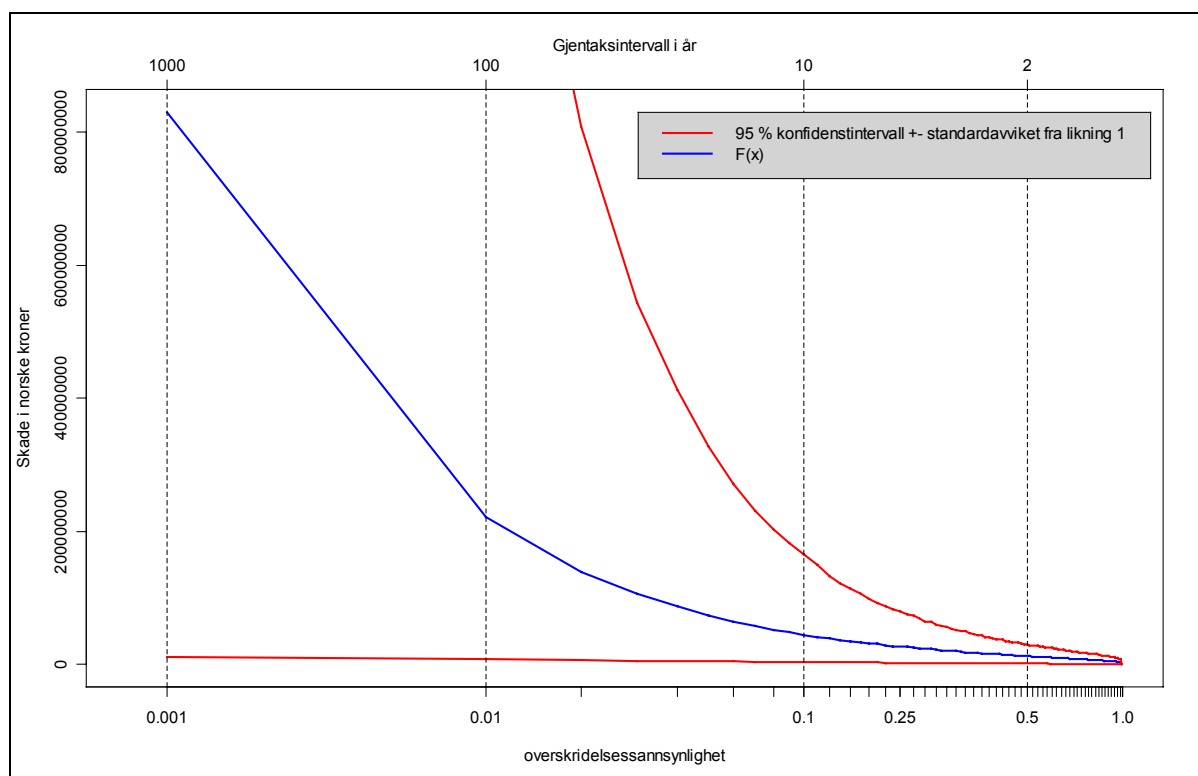
*Figur 25: Til venstre er det en skadekurve for bolighus(111) i Hamar basert på vannstandsdata fra 1909-1961, og figuren til høyre bygger på vannstandsdata fra 1962-2000.*

Det ble i tillegg plottet en kurve for den senere perioden hvor usikkerheten for de individuell huskostnadene ble tatt med i analysen (Figur 27). Det er tydelig å se at her har konfidensbåndet blitt bredere, og nedre konfidensintervall beveger seg så vidt over 0, men det kommer greit fram at det er flomfrekvensanalysen som står for den største delen av usikkerheten.

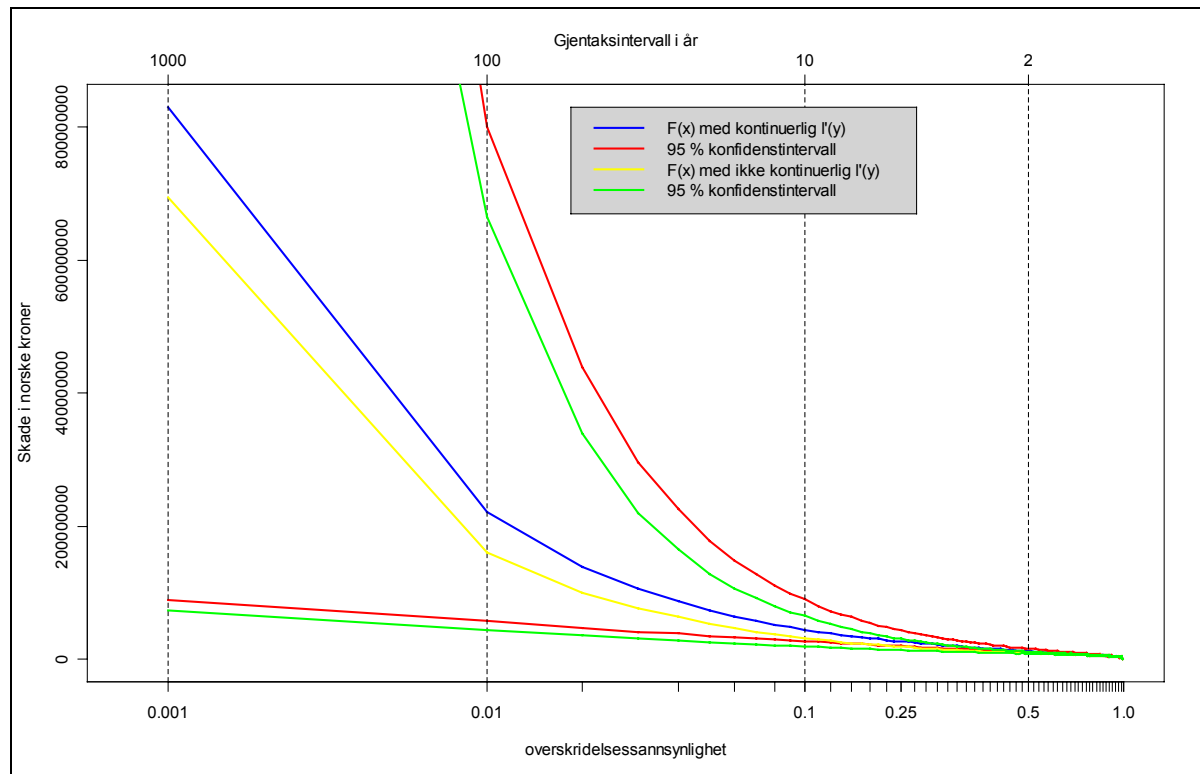
Plottet gir en indikasjon på at det er stor usikkerhet forbundet med slike analyser, og det er viktig at man har så mye data tilgjengelig som mulig. Bare fra Figur 25 er det tydelig å se at kun en dobling av antall data gir et mye tynnere konfidensbånd.



Figur 26: Skadekurve for bolighus(111) i Hamar



Figur 27: Skadekurve for bolighus(111) i Hamar med implementert usikkerhet fra Figur 15.



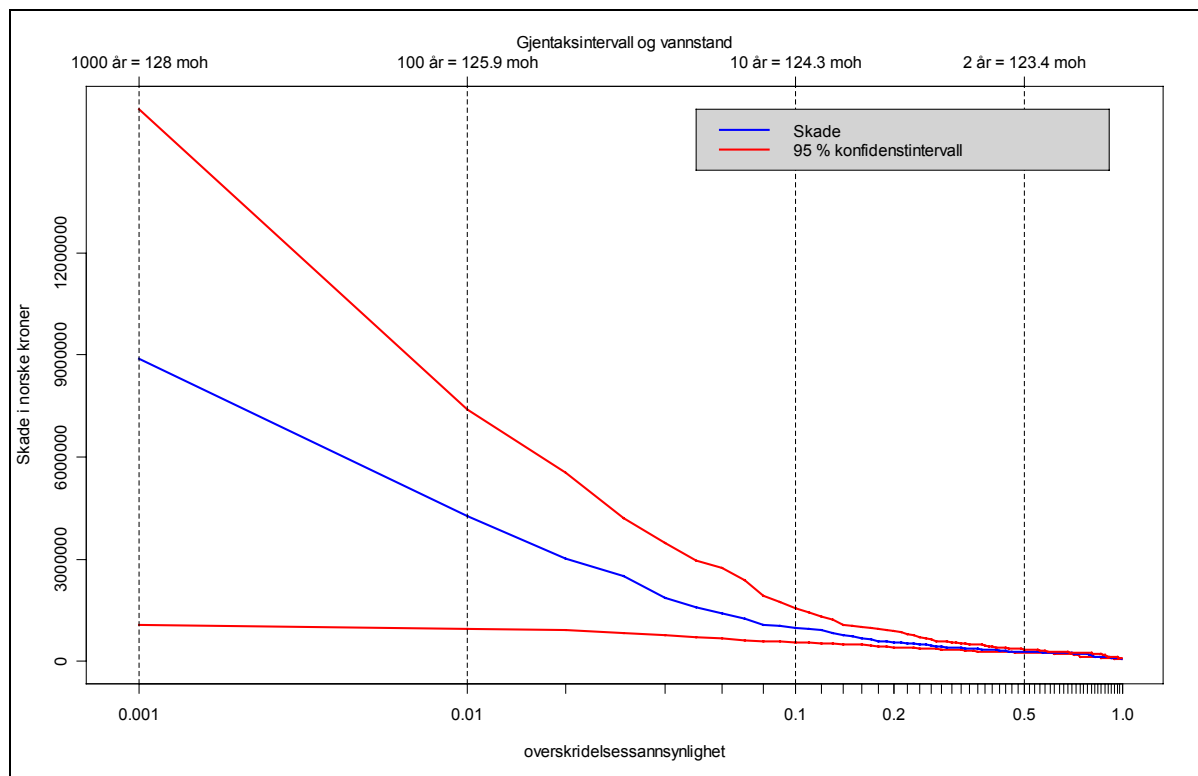
Figur 28: Skadekurve for bolighus i Hamar beregnet både på bakgrunn av kontinuerlige data fra regresjonsanalysen av  $l'(y)$  som er blå likning 1, og høy og liten skade som er gul (likning 59).

Fra Figur 28 kan det observeres at en skadekurve konstruert ved hjelp av likning 1 gir nokså likt resultat som en konstruert med bakgrunn i likning 61, det vil igjen si at det trolig ikke er så farlig hvilke av disse to metodene man velger. Det som er viktig å merke seg er at begge disse metodene er konstruert på bakgrunn av det samme datagrunnlaget, som igjen er veldig usikkert. Men siden de oppfører seg nokså likt, kan det tyde på at det kan være lurt å bruke de kontinuerlige dataene fra likning 61, siden den bygger på mer data. Ulempen med bruk av likning 61 er at skaden til et individuelt hus vil øke kontinuerlig, selv om vannet ligger høyt over selve huset. Det er jo naturlig at skaden hadde blitt konstant ved en bestemt høyde på huset. Siden forskjellen på flomvannstand og normal vannstand ikke er noe særlig større enn 5 meter, og det sjelden er noe mer enn kjelleren som blir berørt, så spiller nok dette en liten rolle.

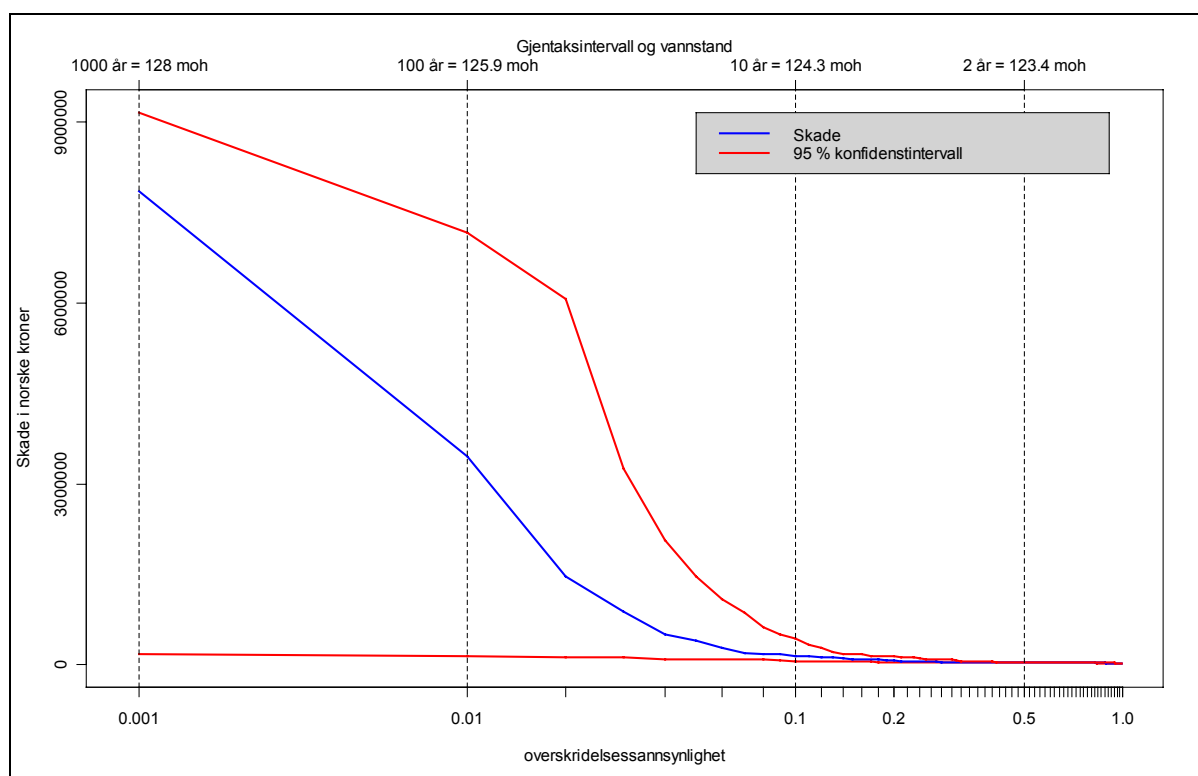
### Skadefunksjon for veier og jernbane

Som grunnlag for beregningene av skadekurver for veier og jernbane er det brukt metodikken beskrevet i kapittel 4.3. Skadekurvene for henholdsvis veier og jernbane er presentert i Figur 29 og Figur 30.





Figur 29: Skadekurve for veier i Hamarområdet



Figur 30: Skadekurve for jernbane i Hamarområdet

Fra Figur 29 og Figur 30 er det tydelig å se at avstanden mellom øvre og nedre konfidensbånd er stor, spesielt når overskridelsessannsynligheten er mindre enn 0,1.

Det er også viktig å notere seg at vannstanden oppgitt i disse plottene er for den blå skadekurven, det blir ikke reelt å se på når man tolker konfidensbåndet. Grunnen til at det nedre konfidensintervallet nesten ikke øker ved lav overskridelsessannsynlighet, skyldes at flomfrekvenskurven til dette intervallet følger en weibullfordeling og vil dermed konvergere mot en vannstand på ca 124,5, som er for lavt til å få høye skadetall. Skadekurven og det øvre konfidensintervallet som begge følger en frechetfordeling, vil aldri konvergere, slik at her vil kostnaden hele tiden øke. Konfidensbåndet kunne trolig blitt smalere hvis det hadde blitt fremstilt globalt. For jernbane kan man se at konfidensbåndet blir smalere for lavere overskridelsessannsynlighet. Det skyldes at det øvre konfidensintervallet er oppe på så høye vannstander at det ikke er så mye mer som kan bli oversvømt.

### *Oppsummering av risikoanalysen for Hamar*

Det er de samme usikkerhetsberegningene som ligger til grunn for skadeplottene i Figur 25, Figur 27 og Figur 29, selv om det er jernbane, hus eller vei. Dermed er det naturlig at usikkerheten oppfører seg likt i de ulike skadekurvene.

Type beregning	Forventet skade på jernbane	Forventet skade på veier	Forventet skade på bolighus
Skade	37 375 NOK	280 800 NOK	13 M NOK
Øvre konfidensintervall	39 000 NOK	360 000 NOK	16 M NOK
Nedre konfidensintervall	37 375 NOK	255 600 NOK	11 M NOK

*Tabell 5.3.1 viser den forventede skaden for bolighus, jernbane og veier i Hamar*

Tallene for forventet skade gitt i Tabell 5.3.1, kan virke litt høye. Det virker ikke helt naturlig at det hvert år skal bli påført skade for 13 millioner på bolighus i Hamar. Selv om Figur 22 viser at det er få bolighus som er utsatt for flom, er det et flertall av de skadeberørte husene som kun blir påført skade i kjellerene. Kjellerhøyden er satt til 2.5 meter, det kan også hende dette er litt høyt. Det er i dette arbeidet ikke undersøkt om troverdigheten til denne verdien.

## 6 Konklusjon

I begynnelsen av dette arbeidet ble det antatt at det var nok kostnadsdata lett tilgjengelig og hovedpoenget var å presentere svarene fra en risikoanalyse for Hamar og Lillestrøm. Etter hvert som det ble oppdaget at det ikke fantes så mye kostnadsdata ble oppgaven mer konsentrert mot beregning av eventuelle usikkerheter i flomfrekvensanalysen. Usikkerhetsberegninger er et lite utbredt tema innenfor hydrologien, og det meste av informasjonen ble hentet fra statistikken.

Beregningen av flomfrekvensanalysen, bød også på en god del problemer da vannstandsdataene dessverre var påvirket av reguleringer opp igjennom årene. Det medførte at det var mest fornuftig å bruke vannføringsdata fra de siste 30-40 årene. For Lillestrøm var det ekstra vanskelig å estimere gode gjentaksintervall. Dette fordi et 100 års gjentaksintervall varierer med opp til 3 meter beregnet ut i fra forskjellige tidsepoker på 1900-tallet. Det er vanskelig å si hva man skal stole på, men om noen år burde det være lettere si noe om dette siden det nå kan se ut som at de største reguleringene er gjort.

Digitale kartgrunnlag var også tilgjengelige, selv om de var noen år gamle. Dataene for Hamar var veldig bra og hadde all den informasjonen som trengs til å utføre en god flomrisikoanalyse. Kartet for Lillestrøm derimot manglet mye god informasjon.

Selv om datagrunnlaget over skadetall var tynt ble det både etablert skadekurve for Hamar og Lillestrøm. For Hamar ble det etablert slike kurver for bolighus, vei og jernbane, men for Lillestrøm ble det kun konstruert for veier og jernbane.

På grunn av en god del reguleringer opp igjennom årene, har man ikke lange nok pålitelige vannstandsserier, slik at usikkerheten blir større enn hva som er optimalt for en risikoanalyse. Videre skaper andre elementer også store usikkerheter slik som kartgrunnlag og skadetall. Skadetallet kan forbedres kraftig ved å foreta en mer inngående analyse av data tilgjengelig etter Vesleofsen (1995), og det kan også til en viss grad være aktuelt å bruke data fra utlandet. Bedre terrengmodeller vil også etter hvert bli mer tilgjengelige. NVE sitt flomsonekartprosjekt krever gode terrengmodeller. Dette vil om noen år være tilgjengelig for både Lillestrøm og Hamar, og for de fleste andre flomutsatte områder i Norge. Det vanskeligste er nok å finne en god skadefunksjon for veier og jernbane. Det er mange faktorer som spiller inn. Blant annet påfører rennende vann mer skade enn stillestående vann. I dette arbeidet bygger analysen kun på at vannet står stille, men skadefunksjonene som er brukt bygger på at det i tillegg er rennende vann tilstede. Som regel vil flommer i sideelver samsvare med flom i hovedvassdraget, slik at tallene blir nokså reelle. Men som kommentar til dette er det verdt å merke seg at veier i byer ikke er på langt nær så utsatt for oversvømmelse fra elver, som veier i skogsområder. Det er ikke foretatt noen beregninger på skade påført veier/jernbane i urbane strøk.

Med de forbedringene nevnt over vil man komme langt på vei med å utarbeide gode skadekurver for de fleste vassdrag i Norge. Med dagens dataprogrammer er det rimelig greit å implementere dette uten altfor mye jobb.

Som en generell konklusjon kan det sies at var veldig vanskelig å gjennomføre en god flomrisikoanalyse for Lillestrøm. I tillegg til at det manglet mye informasjon, ble det ekstra vanskelig på grunn av komplekse reguleringer av vassdraget. I tillegg skal Lillestrøm være sikret mot flom på grunn av omfattende flomvoller og pumpestasjoner. Hamar derimot var mye enklere, den største begrensningen ble her at det var få skadefunksjoner tilgjengelige.

## Tillegg A

### Bygningsfordeling for Hamar, konstruert på bakgrunn av sosi-data

ID	Type bygning	Antall	Antall oversvømt	
			Vesleofsen i 1995	Storofsen i 1789
111	Enebolig	5719	7	75
112	Enebolig m/hybel/sokkelleil.	224	0	0
113	Våningshus	432	0	0
121	Del av tomannsbolig-vertikal	418	7	14
122	Tomannsbolig, horisontaldelt	318	0	2
123	Del av våningh.tomannsb/vert.	2	0	0
124	Våningsh. tomannsb./horisont.	1	0	0
131	Del av rekkeh. m/3-4 boliger	84	0	0
132	Del av rekkeh m/5 bolig el.fl	299	0	0
133	Del av kjede/atr.h innt.4 bol.	130	0	0
134	Del av kj/atr.h m/5 bol. el fl	49	0	0
135	Terrassehus	8	0	0
136	Andre småhus med 3-4 boliger	142	0	3
141	Stort fritt. boligbygg på 2 etasjer.	4	0	0
142	Stort fritt. boligbygg på 3 og 4 etasj.	118	0	1
143	Stort fritt. boligbygg på 5 -> etasjer	18	0	0
144	Stort sammenh.boligbygg på 2 etasjer	0	0	0
145	Stort sammenh.boligbygg på 3 og 4 et	0	0	0
	Stort sammenh.boligbygg på 5 ->			
146	etasjer eller over (5 boliger eller mer)	0	0	0
151	Bo- og servicesenter	1	0	0
152	Studenthjem/studentboliger	11	0	0
159	Annen bygning for bofellesskap /1	1	0	0
	Fritidsbygg(hytter,sommerh. Og			
161	lignende	449	1	3
162	Helårsb.benytted som fritidsb.	0	0	0
163	Våningh. benytted som fritidsb	0	0	0
171	Seterhus, sel, rorbu og lignende	245	1	1
172	Skogs- og utmarkskoie, gamle	13	0	0
181	Garasje ,uthus anneks til bolig	5028	8	57
182	Garasje, uth. anneks til fritidb	125	0	0
183	Naust, båthus, sjøbu	0	0	0
191	Kombinert bolig > annet areal	3	0	0
192	Kombinert bolig < annet areal	6	0	1
193	Boligbrakker	1	0	0
	Annen boligb.(Eks sekundærbolig			
199	reindr.)	21	0	0
211	Fabrikkbygning	24	1	2

212	Verkstedbygning	80	14	21
213	Produksjonshall	31	2	10
214	Bygning for renseanlegg	6	0	0
215	Bygning for avfallshåndtering	2	0	0
216	Bygn. for vannfors. bla. pumpest	26	9	11
219	Annen industribygning /1	32	1	1
221	Kraftstasjon (> 15 000 kVA)	1	0	0
222	Mindre kraftstasjon	0	0	0
223	Transformatorstasj.(>10000 kVA	7	0	0
224	Mindre transform.stasj./kiosk	99	3	5
229	Annen energiforsyningsbygning /1	4	0	0
231	Lagerhall	267	12	29
232	Kjøle- og fryselager	2	0	0
233	Silobygning	25	0	1
239	Annen lagerbygning /1	10	3	6
241	Hus for dyr/landbr.lager/silo	135	0	0
242	Landbruksgarasje/redskapshus	699	0	0
243	Veksthus	34	0	1
244	Driftsb. fiske/fangst/oppdr	1	0	0
245	Naust/redskapshus for fiske	0	0	0
248	Annen fiskeri- og fangstbygn.	2	0	1
249	Annen landbruksbygning	261	2	3
290	Annen industri- og lagerbygn	7	2	2
311	Kontor- og adm.bygning, rådhus	115	5	21
312	Bankbygning, posthus	0	0	0
313	Radio og TV-hus	0	0	0
319	Annen kontorbygning /1	40	2	6
321	Kjøpesenter, varehus	11	2	3
322	Butikk/forretningsbygning	63	3	13
323	Bensinstasjon	20	1	1
329	Annen forretningsbygning /1	11	0	1
330	Messe- og kongressbygning	2	1	1
390	Annen kont. og forretn.bygn	78	0	15
411	Eksp.bygn. flyterm. kontr.tårn	1	0	0
412	Jernbane- og T-banestasjon	2	0	1
413	Rutebilst.,bussterminal,leskur	0	0	0
414	Fergeterminal	0	0	0
415	Godsterminal	0	0	0
416	Postterminal	1	1	1
419	Annen eksp. og terminalbygning /1	10	1	5
421	Telebygning, telefonkiosk	2	0	0
422	Radiolinkstasjon	2	0	0
423	TV-/FM-st.(hovedsenderstasjon)	0	0	0
424	AM-stasjon (kortbølgestasjon)	0	0	0
429	Annen telekommunikasjonsbygn. /1	0	0	0
431	Parkeringshus	8	1	2
432	Bussgar.,trikke-lokomotivstall	2	1	1
433	Flyhangar	4	0	0
439	Annen garasje-/hangarbygning /1	3	0	3
441	Biltilsynsbygning	0	0	0

---

442	Driftssentral for vegvesenet	2	0	0
443	Vakt-/bombygn.tollst., bilvekt	0	0	0
449	Annen veg-og biltilsynsbygning	0	0	0
490	Annen samf. og kom.bygn	3	2	2
511	Hotellbygning	3	0	1
512	Motellbygning	3	0	3
519	Annen hotellbygning /1	0	0	0
521	Hospits, pensjonat	3	0	0
522	Vandre-feriehjem,turisthytte	0	0	0
523	Appartement	0	0	0
524	Camping- /utleiehytte	37	0	2
529	Annen bygning for overnatting /1	5	0	0
531	Restaurantbygning, kafébygning	16	1	3
532	Sentralkjøkken, kantinebygning	0	0	0
533	Gatekjøkken, kioskbygning	3	0	0
539	Annen restaurantbygning /1	0	0	0
590	Annen hotell og rest.bygn	0	0	0
611	Lekepark	0	0	0
612	Barnehage	28	0	0
613	Barneskole	16	0	0
614	Ungdomsskole	6	0	0
615	Kombinert barne- og ungdomsskole	0	0	0
616	Videregående skole	11	0	0
619	Annen skolebygning /1	7	0	0
621	Univ./høgskole m/auditor.leses	4	0	0
622	Spesialbygning	3	0	0
629	Annen universitet/høgskolebygn /1	9	0	0
630	Laboratoriebygning	0	0	0
641	Museum, kunstgalleri	84	8	41
642	Bibliotek, mediatek	3	0	0
643	Zoologisk/botanisk hage (byg.)	0	0	0
649	Annen museum/biblioteksbygning /1	5	0	2
651	Idrettshall	7	0	0
652	Ishall	1	0	0
653	Svømmehall	1	0	0
654	Tribune og idrettsgarderobe	2	0	0
655	Helsestudio	1	0	0
659	Annen idrettsbygning /1	22	0	2
661	Kino/teater/opera/konsertbygn	4	0	0
662	Samfunnshus, grendehus	5	0	0
663	Diskotek	0	0	0
669	Annet kulturhus /1	4	0	0
671	Kirke, kapell	7	0	0
672	Bedehus, menighetshus	9	0	0
673	Kremat., gravkapell, bårhus	2	0	0
674	Synagoge, moske	0	0	0
675	Kloster	0	0	0
679	Annen bygn. for religiøs akt. /1	1	0	0
690	Annen kultur/forskningsbygning	4	0	0
711	Lokalsykehus	0	0	0

---

712	Sentralsykehus	1	0	0
713	Region-, universitetssykehus	0	0	0
714	Spesialsykehus	0	0	0
719	Annet sykehus /1	1	0	0
721	Sykehjem	2	0	0
722	Bo- og behandlingssenter, aldershjem	6	0	0
723	Rehabiliter.institusjon.kurbad	0	0	0
729	Annet sykehjem /1	1	0	0
731	Klinikk, legekantor/legesenter	1	0	0
732	Helse-/sosialsent. helsestasj.	2	0	0
739	Annen primærhelsebygning /1	1	0	0
790	Annen helsebygning	1	0	0
811	Landsfengsel	0	0	0
812	Hjelpefengsel, kretsfengsel	2	0	0
813	Arbeidsskoloni	0	0	0
819	Annen fengselsbygning /1	0	0	0
821	Politistasjon	0	0	0
822	Brannstasjon, ambulansestasjon	1	1	1
823	Fyrstasjon, losstasjon	0	0	0
824	Stasj. for radarov.av fly-/skipstrafikk	0	0	0
825	Tilfluktsrom/bunker	0	0	0
829	Annen beredskapsbygning /1	0	0	0
830	Monument (ny)	0	0	0
840	Offentlig toalett	2	0	0
890	Annet fengsel-/ beredskapsbyg. mv.	1	0	0

*tabell 5.3.2: Tabellen over viser oversikt over andelen av forskjellige bygningstyper i Hamarområdet. Tabellen viser også antallet som hadde blitt utsatt for flom hvis vannstanden hadde vært tilsvarende Stor- og Vesleofsen. Dataene er utarbeidet fra sosi-data, slik at alle bygningstyper representert i sosi er med, selv om de ikke representert i område. ID = BYGTYP\_NBR.*



## Tillegg B

### Arcinfo kommandoer:

Under er det listet opp noen av de viktigste kommandoene brukt i GIS-verktøyet Arcinfo. Sosi-filene er konvertert til Arcinfo sitt utvekslingsformat *e00* i et selvstendig konverteringsprogram. E00-filene er videre importert til Arcinfo som cover, med Import-kommandoen.

**Tegne opp linjer fra coveret lilles på et kart og plukke ut en temakode, og lage et cover nyeveier av denne temakoden:**

```
Arc:      arcedit
Arcedit:  disp 9999
Arcedit:  mapex lilles
Arcedit:  de arc
Arcedit:  edit lilles
Arcedit:  draw
Arcedit:  ef arc
Arcedit:  sel ltema = 7003
Arcedit:  put nyeveier
```

**Lage dtm med navn hamardtm av kurvebildet hoydek:**

```
Arc:      TopoGrid hamardtm 100
TopoGrid: contour hoydek hoyde
TopoGrid: end
```

**Lage grid av et linjecover:**

```
Grid:      Setwindow hamardtm
Grid:      setcell 10
Grid:      veigrd = linegrid(nyeveier)
Grid:      veigrd2 = con(veigrd, hamardtm)
Grid:      veigridet.txt = sample(veigrd2)
```

**tettere mellom punktene i en linje:**

```
densifyarc jernbane jerbaneny 100
```

**Gjøre om linje til punkt:**

```
Arcpoint  jerbaneny banepoint line koordh
```

**Tilordne dtm til punkt:**

```
arc:      Latticespot hamardtm byggpkt moh
```

**Lage punkt av polygon:**

```
Arcedit:  Ef label  
          Sel all  
          Put byggpkt  
Arc:      build byggpkt point
```

**Eksporter data fra arcinfo:****XY-coordinater:**

```
Arc:      ungenerate point hsonel hsonel.txt  
Arc:      Sample (veipunkter,DTM)
```

**Eksportere Tabellinformasjon:**

```
Arc:      tables  
Tables:   Sel hsonel.pat  
Tables:   Unload hsonel.txt
```

---

## KILDELISTE

- Ahmed, A-F. og Stedinger, J.R. (1999): Hydrologic and Economic Uncertainties and Flood-Risk Project Design. *Journal of water resources planning and management*, 125, 314-324.
- Berger, J.O (1985): *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag
- Cheng, R.C.H og Iles, T.C (1983): Confidence Bands for Cumulative Distribution Functions of Continuous Random Variables. *Technometrics* 25, 77-86
- Commission on Geosciences (2000): Risk analysis and uncertainty in flood damage reduction studies. NATIONAL ACADEMY PRESS, Washington D.C.
- Davis, J.C. (2002): *Statistics and data analysis in geology*, third edition. John Wiley & son, Inc.
- Efron, B. og Tibshirani, R. J.(1998): *An introduction to Bootstrap*. CHAPMAN & HALL/CRC
- Gottschalk, L. og Kraskovskaia, I. (1999): Economic Risk of Flooding – a case study for the floodplain upstream Nor in the Glomma River, Norway, HYDRA-rapport nr.R01
- Gottschalk, L og Krasovskaia, I (2001): Regional flood frequency analysis – A theoretical background. *Lecture notes in Stochastic Hydrology*. Universitetet i Oslo.
- Gumbel, E. J. (1958): *Statistics of Extreme*. Colombia Univ. Press, New York
- Hosking, J.R.M. og Wallis, J.R (1997): *Regional Frequency Analysis*.
- Hutchinson, M.F(1989): A new procedure of gridding elevation and stream line data with automatic removal of spurious pits, *Journal of Hydrology*,106,211-232
- Jeng, S-L og Meeker, W.Q. (2001): Simultaneous Parametric Confidence Bands for Cumulative Distributions from Censored Data. *Technometrics* 43, 450-461
- Kotz, S. og Nadarajah, S. (2000): *Extreme Value Distributions: Theory and Applications*. Imperial College Press.
- Krzysztofowicz, Roman (1983): Category-Unit Loss Functions for Flood Forecast-Response System Evaluation. *Water Resources Research*, 19(6), 1476-1480.
- Lamber, J.H og Li, D. (1994): Evaluating Risk of Extreme Events For Univariate-Loss Functions. *Journal of Water Resources Planning and Management*,123(3),382-399.
- Larsen, R. J. og Marx, M. L (1986): *An introduction to MATHEMATICAL STATISTICS and its APPLICATIONS*, Prentice-Hall.

- Liestøl, K og Storvik, G: Stokastisk simulering, bootstrapping og Monte Carlo testing, Universitetet i Oslo 2001
- Mitsiopoulos, J., Haimes, Y.Y. og Li, D. (1991): Approximating Catastrophic risk through statistics of extremes. *Water Resources Research*, 27(6), 1223-1230
- Penning-Rowsell, E.C. og Chatterton, J.B. (1977): The benefits of flood alleviation. Gower Publishing Company.
- Petterson, L-E (1997) : Hydrologisk data for Mjøsa, NVE Publikasjon nr. 6.
- Petterson, L-E (2000) : Flomberegning for Mjøsa og Vorma, NVE Publikasjon nr 21.
- Petterson, L-E (2002): Flomberegning for Nedre Glomma, NVE Publikasjon nr 15 .
- Rao, A.R. og Hamed, K.H. (2000): Flood Frequency analysis,. CRC Press.
- Rice, J.A. (1995): Mathematical Statistics and Data Analysis, Second Edition. Duxbury press.
- Statens kartverk(1987): Sosi-standard: <http://www.statkart.no/standard/sosi/html/sosi.htm>
- Sælthun, N.R. m. flere (2000): Økonomisk risikoanalyse for flommer, HYDRA-rapport nr. R03.
- Venables, W.N og Ripley, B.D(1999): Moderen Applied Statistics with S-PLUS, Third Edition. Springer-Verlag.
- Wathne, M. m. flere (1999): Samfunnskostnader på grunn av flom i vassdrag. HYDRA-rapport nr. R02.

## Stikkordregister

Anderson-Darling test, 38

Annual maximum, 16, 22, 30, 34, 35, 39, 61

Arcinfo, GIS-verktøy, 7, 9, 10, 33, 45, 64

Aursjø, 11

Bias, 22, 23, 24, 42, 43

Bootstrapping, 22

Bygninger, 9, 50

Enebolig, 10, 60

Flomfrekvensanalyse, 4, 5, 16, 30

Forventet skade, 57

Frechetfordeling, 18, 19, 20, 42

Generalized extreme value fordeling, 18, 19, 24, 41

Geografiske informasjonssystemer, 8

GEV-fordeling, 15, 18, 19, 20, 23, 30, 37, 38, 41, 44

Gjentaksintervall, 21

Glomma, Norges lengste elv, 11, 34, 39, 41, 47, 66, 67

Grid, 64

Gumbelfordeling, 18, 19, 20, 37, 39, 44, 66

Hamar, 1, 2, 5, 6, 8, 10, 11, 33, 34, 35, 36, 37, 38, 41, 45, 47, 49, 50, 53, 54, 55, 56,  
57, 58, 59, 60

Jernbane, 13, 61

Kjernetetthetsestimering, 16

Konfidensbånd, 25, 26, 43, 44, 46, 53, 57

Konfidensintervall, 23, 24, 25, 27, 42, 43, 53, 57

Kumulativ fordelingsfunksjon, 14

Lågen, 12

Leira, 11, 47

Vesleofsen, 1995, 58, 60, 63

Lillestrøm, 1, 2, 5, 6, 8, 11, 12, 33, 34, 41, 45, 46, 47, 48, 49, 58, 59

L-momenter, 4, 17, 18, 23, 43, 46

Maximum likelihood, 42, 43

Mjøsa, 2, 4, 5, 6, 7, 10, 34, 35, 36, 37, 39, 42, 44, 47, 50, 67

Mørkfoss, 10, 11, 34, 35, 38, 44

Nittelva, 11, 47

Norges vassdrags- og energidirektorat, 6, 8, 10, 58, 67

Østmarka, 11

Øyeren, 2, 4, 5, 6, 7, 10, 11, 12, 35, 37, 38, 39, 42, 43

Peek over treshold, 16, 22, 30, 34, 35

Probability weighed moments, 17

Regional flomfrekvensanalyse, 39

Resampling, 24

Risiko, 26

Risikoanalyse, 6, 46

Sannsynlighetstetthet, 14

Simulering, 23

Skade, 12, 57

Skadefunksjonen, 26, 33, 50

Skedsmo kommune, 6, 8

Solbergfoss, 12

SOSI, norsk kartstandard, 9, 50

Statens kartverk, 8, 10, 67

Storofsen, 1789, 46, 60

Svanfoss, 10, 37

Usikkerhet, 40, 44

Weibullfordeling, 18, 19, 20, 42, 43